

Rui Miguel Castro Ramos

Determinação computacional dos efeitos da mutagénese em interfaces proteína-DNA



DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO
SETEMBRO/2012

Rui Miguel Castro Ramos

Determinação computacional dos efeitos da mutagénese em interfaces proteína-DNA



Tese submetida à Faculdade de Ciências da Universidade do Porto para a obtenção do grau de Mestre em Química

Composição do Júri:

*Professor Doutor Alexandre Lopes Magalhães
Professor Doutor José Ferreira Gomes
Professora Doutora Maria João Ramos
Doutor Sérgio Filipe Sousa*

DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO
SETEMBRO/2012

Agradecimentos

À Professora Doutora Maria João Ramos e ao Professor Doutor Pedro Alexandrino Fernandes pela possibilidade que me deram de trabalhar no grupo, de descobrir e fomentar o gosto por esta área da Química e pelos ensinamentos prestados no decorrer deste trabalho.

À Doutora Irina Sousa Moreira que me acolheu desde o primeiro dia e me ensinou grande parte do que hoje sei sobre Química Teórica. Por ter estado sempre disponível quando precisava, pelo apoio, ajuda e motivação que me deu.

A todos os meus colegas do grupo pela simpatia e amizade com que me acolheram. Em particular a todos aqueles que estiveram no laboratório 3.26, à Elizabeth, à Bruna, à Ana, ao Marco e César pelos bons momentos que ali passamos todos os dias.

Aos meus amigos, por estarem sempre presentes e me aturarem nos dias em que as coisas corriam menos bem. Particularmente ao João, com o qual passei grande parte dos meus dias nos últimos anos.

À minha família, por tudo o que sempre me proporcionaram e por toda a motivação que me deram para atingir os meus objetivos. Sem eles nada teria sido possível.

Finalmente, à Elsa. Por ter estado sempre presente, nos bons e nos momentos menos bons, por todo o apoio e compreensão e por aquilo que significa para mim.

Resumo

A descrição das propriedades funcionais e estruturais presentes em interações proteína-proteína, proteína-ligando e proteína-DNA é fundamental e de grande interesse para a comunidade científica, uma vez que os complexos formados estão presentes na grande maioria dos processos biológicos. Neste trabalho pretendeu-se estudar um conjunto de interfaces proteína-DNA e suas características, recorrendo para isso a diversas técnicas computacionais tais como: simulações de dinâmica molecular (DM) e aplicação da metodologia de mutagénese sequencial da interface (*Alanine Scanning Mutagenesis*) para identificação de *hot-spots* e *null-spots*. Estudou-se, também, um dos mais importantes complexos proteína-DNA: um sistema formado pela enzima Transcriptase Reversa (HIV-1 RT) complexada com uma cadeia dupla de DNA e com um inibidor do vírus HIV-1, Tenofovir (TDF).

Analisaram-se cinco complexos proteína-DNA, para os quais se realizaram simulações de DM de 10 ns em solvente explícito e implícito, utilizando o campo de forças *ff99SB* do Amber. Um total de 60 resíduos, dos 5 complexos, foram mutados para alanina e a sua energia livre de ligação calculada e comparada com os valores experimentais. Depois de se terem avaliado diferentes parâmetros como a influência do tempo de simulação, representação do solvente, número de trajetórias, força iónica e constante dielétrica conseguiu-se obter uma boa correlação com os valores experimentais, com um erro médio de 2,86 kcal / mol e 60 % de especificidade do método. A taxa de sucesso obtida é ainda mais relevante uma vez que não existem estudos semelhantes para este tipo de interface. Este trabalho permitirá a abertura de novas perspetivas para futuros estudos em interfaces proteína-DNA, com a determinação dos resíduos mais importantes para a ligação, tal como tem sido efetuado para interfaces proteína-proteína.

Estudou-se também em maior detalhe um sistema proteína-DNA de grande importância, a HIV-1 RT, que é uma das enzimas responsáveis pela replicação do vírus HIV-1. Apesar de ser um dos inibidores do HIV-1 mais utilizados, ainda não se conhece completamente o mecanismo de excisão do TDF quando este se encontra incorporado pela HIV-1 RT. Realizaram-se simulações de DM dos complexos da HIV-1 RT, com o TDF incorporado no P-site e no N-site, para aferir os fatores que estavam por trás da menor excisão do TDF. Os resultados obtidos parecem indicar que a flexibilidade do TDF no P-site dando origem a várias conformações será o fator preponderante para se obter uma menor excisão.

Abstract

The description of the structural and functional properties of protein-protein, protein-ligand and protein-DNA interactions is important and of great interest for the scientific community, because these complexes are present in most biological processes. In this work we intended to study a set of protein-DNA interfaces and their characteristics, using several computational techniques such as: molecular dynamics simulations (MD) and the application of the Alanine Scanning Mutagenesis methodology for the identification of hot-spots and null-spots. We have also studied one of the most important protein-DNA complexes: a system formed by the HIV-1 Reverse Transcriptase (HIV-1 RT), double stranded DNA and the HIV-1 inhibitor Tenofovir (TDF).

We analyzed five protein-DNA complexes for which MD simulations were carried out for 10 ns in explicit and implicit solvent, using the Amber force field *ff99SB*. A total of 60 residues, from the five complexes, were mutated to alanine and their binding free energy calculated and compared to the experimental values. After having evaluated the influence of the different parameters such as the simulation time, representation of the solvent, number of trajectories, ionic strength and dielectric constant, we were able to get a good correlation with the experimental values with an average error of 2,86 kcal / mol and 60 % specificity of the method. The success rate achieved is relevant since there are no similar studies for this type of interface. This work will open new perspectives for future studies on protein-DNA interfaces, with the determination of the most important residues for binding, as it was done for protein-protein interfaces.

We have also studied in greater detail a very important protein-DNA system, the HIV-1 RT, which is one of the enzymes responsible for HIV-1 replication. Despite being one of the most prescribed HIV-1 inhibitors, the excision mechanism of TDF when it is incorporated into the HIV-1 RT is still unknown. We conducted MD simulations of the HIV-1 RT complexes with TDF at the P-site and N-site to access the factors behind the smaller excision rate of TDF. Our results seem to indicate that the flexibility of TDF at the P-site, and its multiple conformations, could be the dominant factor behind its lower excision rate.

Résumé

La description des propriétés fonctionnelles et structurales présentes dans des protéines-protéines, protéines-DNA est d'un grand intérêt pour la communauté scientifique, car les complexes formés sont présents dans la plupart des procès biologiques. Avec cette recherche, on a essayé étudier un ensemble de protéine-DNA interfaces et ses caractéristiques, en recourant à plusieurs techniques informatiques telles que : simulations de dynamique moléculaire (DM) et application de la méthodologie ASM pour l'identification de *hot-spots* et *null-spots*. On a étudié, aussi, un des plus important complexes protéine-DNA : un système formé par l'enzyme HIV-1 RT complexée par une chaîne double de DNA et avec un inhibiteur du virus HIV-1, Tenofovir (TDF).

Cinq complexes protéine-DNA ont été analysés, pour lesquelles ont été réalisées simulations de DM de 10 ns en champ de force de Amber (*ff99SB*). Un total de 60 résidus, de 5 complexes, ont été changes en alanine et son énergie libre de liaison calculée et comparée avec les valeurs expérimentales. Après avoir évalué différents paramètres avec l'influence du temps de simulation, représentation du solvant, nombre de trajectoires, la force ionique et constante diélectrique on a réussi une bonne corrélation avec les valeurs expérimentales, avec un erreur médium de 2,86 kcal / mol et 60 % spécificité du méthode. La taxe de succès obtenue est encore plus relevé car il n'y a pas d'études pareils à propos de ce type d'interface, ce travail permettra l'ouverture de nouvelles perspectives pour de future études en interfaces protéines-DNA, avec la détermination des résidus plus importants pour la liaison, comme a vient d'être effectué pour interfaces protéine-protéine.

A également été étudiée plus en détail un système protéine-DNA d'une grande importance, le HIV-1 RT, qui est une enzyme responsable pour la réplication du HIV-1. Tout en étant un inhibiteur de HIV-1 le plus utilisé, mais pas complètement connue le mécanisme d'excision de TDF quand il est incorporé dans la HIV-1 RT. Ont effectué des simulations DM des complexes de HIV-1 RT, avec TDF incorporés dans le P-site et N-site pour évaluer les facteurs qui étaient derrière la partie inférieure excision de TDF. Les résultats obtenus semblent indiquer que la flexibilité de TDF donnant lieu à différentes conformations est le facteur prédominant d'obtenir une excision inférieure.

Índice

Agradecimentos.....	V
Resumo.....	VII
Abstract	IX
Résumé.....	XI
Índice	XIII
Índice de Figuras	XV
Índice de Tabelas.....	XIX
Lista de Abreviaturas	XXIII
Capítulo I. Introdução	1
1. Sistemas Biológicos.....	3
1.1. Sistemas proteína-proteína.....	3
1.2. Sistemas proteína-DNA	5
2. Um caso particular de um sistema proteína-DNA: HIV-1 RT	9
2.1. HIV	9
2.1.1. Transcriptase Reversa e o vírus HIV-1	9
2.1.1.1 Constituição e domínios da HIV-1 RT.....	10
2.1.1.2 Mecanismo de inibição / resistência da HIV-1 RT	11
2.1.2. Tipos de Inibidores da HIV-1 RT	12
2.1.3. Tenofovir.....	14
2.1.4. Mutações que conferem resistência à HIV-1 RT	14
3. Metodologia Computacional	16
3.1. Mecânica Molecular e Campos de Força.....	17
3.2. Minimização da Energia	23
3.3. Dinâmica Molecular.....	24
3.3.1. Ensembles e Equações do Movimento.....	25
3.3.2. Escolha do tempo de integração.....	26
3.3.3. Condições Periódicas de Fronteira.....	27
3.3.4. Interações de curto e longo alcance.....	28
3.3.5. Controlo da Temperatura e Pressão	29
3.4. Método de cálculo da energia livre	29
3.4.1. MM-PBSA	30
3.4.2. Solvatação	31
3.4.3. Constante dieletrica interna.....	33

3.5. Método computacional Alanine-Scanning Mutagenesis	34
Capítulo II. Metodologia	37
1. MUTAGÊNESE EM INTERFACES PROTEÍNA-DNA	39
1.1. Preparação dos sistemas.....	39
1.2. Simulação de Dinâmica Molecular	42
1.3. Alanine-Scanning Mutagenesis	42
1.4. Análise dos resultados.....	45
1.4.1. RMSD.....	45
1.4.2. RDF	45
2. MECANISMO DE INIBIÇÃO DO TENOFOVIR	46
2.1. Preparação dos sistemas.....	46
2.2. Simulações de Dinâmica Molecular	48
2.3. Análise dos resultados.....	49
Capítulo III. Resultados e Discussão.....	51
1. MUTAGÊNESE EM INTERFACES PROTEÍNA-DNA	53
1.1. Composição das interfaces.....	53
1.2. Estabilidade dos sistemas proteína-DNA: cálculo de RMSD.....	53
1.3. Importância do tempo de simulação na determinação de $\Delta\Delta G$	55
1.4. Importância do tipo de solvente na determinação de $\Delta\Delta G$	59
1.5. Influência do número de DM no cálculo de $\Delta\Delta G$	62
1.6. Importância da força iônica na determinação de $\Delta\Delta G$	65
1.7. Importância da constante dielétrica na determinação de $\Delta\Delta G$	71
1.8. RDFs	75
1.9. SASAs.....	79
2. MECANISMO DE INIBIÇÃO DO TENOFOVIR	81
2.1. RMSD	82
2.2. Estudo da flexibilidade do TDF no P-site.....	85
2.2.1. Medição de ângulos diedros	85
2.2.2. Cálculo de distâncias entre bases emparelhadas	90
2.2.3. Efeito da flexibilidade no mecanismo de excisão	98
2.3. Comparação entre a mutação K65R e a estrutura Wild-Type	100
Capítulo IV. Conclusão e Trabalho Futuro	105
Capítulo V. Referências	109
Capítulo VI. Informação Suplementar	117

Índice de Figuras

Figura I.1 - Representação de HS (vermelho) e NS (amarelo) numa interface proteína-proteína (PdbID: 1FCC [10]).	5
Figura I.2 - Representação dos vários tipos de Fatores de Transcrição. Proteína está representada em cartoon e o DNA em vdw.	7
Figura I.3 - Representação da HIV-1 RT (PDBID:1T05). <i>Fingers, palm, thumb, connection</i> e a RNase H da subunidade p66 estão representados a azul, vermelho, rosa, ciano e cinzento, respetivamente. A subunidade p51 encontra-se representada a castanho. O template e o primer de DNA estão representados a amarelo e laranja.	10
Figura I.4 - Modelo simples representativo do equilíbrio entre o processo de translocação e de excisão. A inibição da síntese de DNA, devido à formação de um complexo terminal, é favorecida quando o primer se encontra ligado no P-site. Adaptado (Menéndez-Arias, L. [25]).	12
Figura I.5 - Estrutura química de vários NRTIs e NNRTIs.	13
Figura I.6 - Ilustração do oscilador harmónico.	20
Figura I.7 - Representação gráfica da forma dos potenciais de Morse e Harmónico.	21
Figura I.8 – Representação da variação da energia torsional para a molécula de etano, como resultado da rotação do ângulo diedro. O máximo corresponde à conformação eclipsada e o mínimo à conformação alternada.	22
Figura I.9 – Representação gráfica de um potencial de Lennard-Jones 12-6. ϵ é a energia mínima da função e R a distância na qual o mínimo da curva ocorre. Para distâncias interatómicas grandes existe uma tendência para zero e para curtas distâncias repulsões muito acentuadas.	23
Figura I.10 - Representação esquemática de condições periódicas de fronteira. Em realce, no centro, encontra-se a célula de simulação e à sua volta as células réplica com o mesmo tipo de movimento da célula central.	28
Figura I.11 - Ciclo termodinâmico utilizado no cálculo da energia livre do complexo.	30
Figura I.12 – Resumo da metodologia computacional ASM aplicada ao estudo de sistemas proteína-proteína	35
Figura II.1 - Representação de um sistema solvatado com uma caixa de águas do tipo TIP3P.	40
Figura II.2 - Representação dos cinco complexos proteína-DNA estudados neste trabalho.	41
Figura III.1 - Representação da variação de RMSD da componente proteica e DNA para os sistemas 1B3T (a), 1BDT (b), 1MNM (c), 1MSE (d) e 1QRV (e).	55

Figura III.2 - Representação da distribuição dos valores de $\Delta\Delta G_{MM-PBSA}$ (kcal mol ⁻¹) para um conjunto de 25 (2 ns) e 75 estruturas (6 ns).	57
Figura III.3 - Representação esquemática das diferenças entre os dois métodos.	63
Figura III.4 - Representação esquemática da metodologia computacional ASM proposta para interfaces proteína-DNA.	75
Figura III.5 - Representação do perfil de RDF e número médio de águas à volta de um HS (a, c) e de um NS (b, d).	76
Figura III.6 - Representação dos valores médios de moléculas de água em redor de HS e NS para distâncias de 3 Å (a), 4 Å (b), 5 Å (c) e 6 Å (d), para os cinco complexos estudados.....	78
Figura III.7 - Representação esquemática dos valores de $\Delta SASA$ e $relSASA$ obtidos para as 60 mutações analisadas.	80
Figura III.8 - Representação do centro activo do complexo HIV-1 RT / DNA / TDF, no P-site (PDB ID: 1T03). TDF na conformação A (verde), TDF na conformação B (magenta), primer (amarelo), template (azul).	82
Figura III.9 - Representação da variação de RMSD da componente proteica, DNA e ligando para os sistemas: a) 1T03A; b) 1T03B; c) 1R0A; d) 1T05; e) 1RTD; f) 3JSM; g) 3JYT; h) 1N5Y	84
Figura III.10 - Representação de alguns dos ângulos diedros medidos neste trabalho: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3). TDF e dCMP estão coloridos a verde e amarelo, respectivamente.	86
Figura III.11 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1T03A: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3).	87
Figura III.12 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1T03B: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3).	88
Figura III.13 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1R0A.	89
Figura III.14 - Representação da conformação inicial das estruturas cristalográficas 1T03A (verde) e 1T03B (azul). Na estrutura 1T03B o TDF encontra-se numa posição em que a sua base está praticamente paralela à base do resíduo dCMP (magenta).	91
Figura III.15 - Representação da distância medida entre as bases adenina (TDF) e citosina (dCMP) para os complexos 1T03A (a), 1T03B (b) e 1R0A (c).	92

Figura III.16 - Representação das conformações adotadas pelo TDF na estrutura 1T03A no P-site (amarelo). As estruturas cristalográficas 1T03A e 1T03B estão representadas a azul e vermelho, respetivamente. A) Conformação anti-paralela; B) Conformação paralela; C) Conformação intermédia.	94
Figura III.17 - Representação das conformações adotadas pelo TDF partindo da estrutura 1T03B (amarelo). As estruturas cristalográficas 1T03A e 1T03B estão representadas a azul e vermelho, respetivamente. A) Conformação 0-2 ns; B) Conformação 2-6 ns; C) Conformação 6-10 ns; D) Conformação 10-18 ns.	96
Figura III.18 - Representação das conformações iniciais das estruturas 1T03A (verde), 1T03B (azul) e 1R0A (vermelho).	97
Figura III.19 – RMSD entre as bases (adenina e citosina) da estrutura 1T03A (a) e 1T03B (b).	98
Figura III.20 - Representação do AZT (magenta) no centro ativo, primer (azul) e template (amarelo).	99
Figura III.21 - Representação da interação entre os resíduos Arg65, Arg72 e Ligando na estrutura 3JSM (a), 1T05 (b), 3JYT (c) e 1RTD (d).	101
Figura III.22 - Representação da orientação conformacional do TDF, Arg72 e Arg 65 na simulação de DM da estrutura 1T05. Imagem retirada aos 4 ns.	103

Índice de Tabelas

Tabela II.1 - Composição em termos de resíduos (número de aminoácidos, ácidos nucleicos, águas e iões) e átomos dos sistemas proteína-DNA em estudo.	40
Tabela II.2 - Resumo das mutações em estudo e seus valores de energia livre experimentais.	43
Tabela II.3 - Descrição dos vários sistemas em estudo. Informação do tipo de ligando e o seu local de ligação no complexo, bem como a resolução da estrutura cristalográfica.	47
Tabela II.4 - Descrição da constituição dos vários sistemas submetidos a DM. Apresenta-se a sua constituição em aminoácidos, ácidos nucleicos, moléculas de água e iões.	48
Tabela III.1 – Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns e 6 ns em solvente explícito. Constantes dielétricas 3 e 7.	55
Tabela III.2 – Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ para a simulação de 2 ns.	58
Tabela III.3 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ para a simulação de 6 ns.	58
Tabela III.4 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e em solvente implícito. Constantes dielétricas 3 e 9.	59
Tabela III.5 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ utilizando solvente implícito.	61
Tabela III.6 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e utilizando outra alternativa para o cálculo da energia livre de ligação, utilizando duas trajetórias. Constantes dielétricas 3 e 9.	63
Tabela III.7 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ utilizando várias trajetórias.	65
Tabela III.8 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação linear e não-linear de Poisson-Boltzmann. Constantes dielétricas 3 e 9.	66
Tabela III.9 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ pela resolução da equação não-linear de PB.	68

Tabela III.10 - Comparação entre os valores obtidos para uma mutação utilizando concentrações iônicas de 0,1 e 0,01M.....	68
Tabela III.11 - $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$ calculado pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação não-linear de Poisson-Boltzmann.....	69
Tabela III.12 - Resultados do erro associado ao cálculo de $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$ pela resolução da equação não-linear de PB.	70
Tabela III.13 - Resultados globais em função do melhor método e constante dielétrica, para os grupos de resíduos carregados, polares e apolares.	72
Tabela III.14 - Resultado do erro médio e dos diferentes testes estatísticos para os resultados obtidos por aplicação dos métodos referidos anteriormente.	74
Tabela III.15 - Número médio de moléculas de água localizadas a distâncias compreendidas entre 3 e 6 Å dos HS e NS, para cada um dos complexos estudados.....	77
Tabela III.16 - Número médio de águas localizadas a distâncias compreendidas entre 3 e 6 Å para o conjunto de HS e NS em estudo.	77
Tabela III.17 - Resultados da distância média e tempo de ocupação entre o Ligando e a Lys/Arg65 e Arg72.....	102
Tabela VI.1 – Resultados da energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns e 6 ns em solvente explícito. Constantes dielétricas de 1 a 10.	119
Tabela VI.2 – Resultados da energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e em solvente implícito. Constantes dielétricas de 1 a 10.....	121
Tabela VI.3 – Resultados da energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, e utilizando outra alternativa para o cálculo da energia livre de ligação, utilizando duas trajetórias. Constantes dielétricas de 1 a 10.	123
Tabela VI.4 – Resultados da energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação linear e não-linear de Poisson-Boltzmann. Constantes dielétricas de 1 a 10.	127
Tabela VI.5 - Resultados dos testes estatísticos para as 60 mutações.....	129
Tabela VI.6 - Resultados dos testes estatísticos para os resíduos carregados.....	129
Tabela VI.7 - Resultados dos testes estatísticos para os resíduos polares.....	130

Tabela VI.8 - Resultados dos testes estatísticos para os resíduos apolares.....	131
---	-----

Lista de Abreviaturas

AMBER – Assisted Model Building with Energy Refinement
ASA – Accessible Solvent Area
ASM – Alanine Scanning Mutagenesis
CF – Campo de Força
CFF – Consistent Force Field
CHARMM – Chemistry at Harvard Macromolecular Mechanics
CPF – Condições Periódicas de Fronteira
DM – Dinâmica Molecular
GAFF - General Amber Force Field
GB – Generalized Born
GROMOS – Groningen Molecular Simulation packages
HF – Hartree-Fock
HIV – Human Immunodeficiency Virus (ou virus da imunodeficiência humana, VIH)
HS – Hot-spot
IN – Integrase
MM – Mecânica Molecular
MM2/3/4 – Molecular Mechanic Force Field for Small Molecules
MQ – Mecânica Quântica
NNRTI – Non-nucleoside Reverse Transcriptase Inhibitors
NPT – Ensemble isotérmico-isobárico
NRTI – Nucleoside Reverse Transcriptase Inhibitors
NtRTI – Nucleotide Reverse Transcriptase Inhibitors
NVE – Ensemble microcanónico
NVT – Ensemble canónico
NS – Null-spot
OPLS – Optimized Parameters for Liquid Simulation
PB – Poisson-Boltzmann
PME - Particle Mesh Ewald
PR – Protease

RESP – Restrained ElectroStatic Potencial

RMSD – Root Mean Square deviation

RT – Reverse Transcriptase

SASA – Solvent Accessible Surface Area

SANDER - Simulating Annealing with NMR-Derived Energy Restraints

TDF – Tenofovir Disoproxil Fumarate

TIP3P – Modelo de adição de solvente

VMD - Visual Molecular Dynamics

μ NV – Ensemble grande canónico

Capítulo I. Introdução

CAPÍTULO I: INTRODUÇÃO

1. Sistemas Biológicos

A natureza é constituída por um grande número de sistemas biológicos, tanto à macro como à microescala, cujas interações são a base de uma quantidade infindável de processos. As proteínas são um dos elementos fundamentais. Estas macromoléculas desempenham funções cruciais em essencialmente todos os processos biológicos, atuando como catalisadores, transportadores e armazenadores: fornecem apoio mecânico e proteção imunitária, transmitem impulsos nervosos, controlam o crescimento e diferenciação, entre outras funções. Outra classe muito importante são os ácidos nucleicos. Os ácidos nucleicos incluem o DNA e o RNA e são os responsáveis pela codificação, transmissão e expressão da informação genética. As proteínas têm a tendência de se associarem com outras macromoléculas, formando complexos estáveis que são a base de muitas funções celulares. Para tal são essenciais as interações proteína-proteína, proteína-ligando e proteína-ácido nucleico. A correta compreensão deste tipo de interações, devido à sua importância, tem vindo a despertar grande interesse na comunidade científica. Neste trabalho, iremos focar-nos, essencialmente, nas interações proteína-DNA.

1.1. Sistemas proteína-proteína

A tendência das proteínas para se ligarem umas às outras, de uma forma altamente específica, formando complexos estáveis é fundamental para os processos biológicos. Estudos baseados em estruturas cristalográficas de proteínas co-cristalizadas com vários ligandos e estudos mutagénicos sequenciais da interface (*Alanine Scanning Mutagenesis*, ASM) permitem obter conhecimento sobre os princípios que governam a formação dos complexos[1]. O método de ASM tornou possível a deteção de epítomos funcionais (resíduos em contacto com um ligando) e demonstrou que a maior parte da energia de ligação proteína-proteína está relacionada com um grupo de alguns aminoácidos localizados na interface da proteína: os hot-spots (HS). A identificação de HS, que são considerados como os resíduos mais importantes para a formação do complexo e para a sua estabilidade, é uma ferramenta deveras importante para a correta compreensão dos princípios que regem a formação dos complexos[2]. Este tipo de interface tem sido extensivamente estudada nos últimos anos. Na atualidade é aceite que as interfaces proteína-proteína são formadas por proteínas com forma e propriedades eletrostáticas adequadas, com um tamanho médio de 1200-2000 Å² [3]. Estas interfaces são

Capítulo I: Introdução

frequentemente hidrofóbicas, sendo que estas interações ocorrem entre regiões apolares dos seus aminoácidos, através de interações de van der Waals, conduzidas pelo ganho de energia livre resultante do movimento de um ambiente polar (aquoso) para um apolar. As interações eletrostáticas são as outras forças de destaque envolvidas em interações proteína-proteína, uma vez que a complementaridade eletrostática de superfícies proteicas promove a formação do complexo. Com a aplicação da metodologia ASM, e consequente detecção dos HS, ficou clara a existência de uma distribuição desigual da contribuição energética de resíduos individuais ao longo da interface, e que apenas alguns contribuem significativamente para a energia livre de ligação de complexos proteína-proteína. Os HS foram definidos como resíduos nos quais a mutação para alanina causa um aumento na energia livre de ligação de, pelo menos, 2,0 kcal/mol [4]. Estes resíduos formam regiões compactas e centrais, cruciais para a associação entre proteínas. Raramente são encontrados HS na extremidade de uma interface [5]. Relativamente à composição em aminoácidos várias análises demonstraram uma distribuição irregular, sendo os fundamentais: triptofano, arginina e tirosina. Por outro lado, aminoácidos como a leucina, serina, treonina ou valina não são geralmente associados a HS, apesar da sua importância em várias estruturas proteicas. O triptofano é o HS mais representativo, desempenhando um papel importante devido ao seu tamanho e natureza aromática, contribuindo com interações π , atuando como dador de hidrogénio e oferecendo proteção a ligações de hidrogénio mais fracas da ação da água, devido à sua elevada superfície hidrofóbica. A arginina é capaz de formar até cinco pontes de hidrogénio e uma ligação iónica com a sua carga positiva [2, 6]. A importância dos resíduos aromáticos para a interação entre proteínas tem sido alvo de diferentes estudos [7]. A maioria destas interfaces são caracterizadas por *pockets* complementares na região central da interface, enriquecidos com resíduos conservados, com elevada complementaridade tanto em tamanho como na justaposição entre HS hidrofóbicos e hidrofílicos. Geralmente, o HS de uma face interatua com um HS da outra face, estabelecendo uma região determinante para a ligação do complexo [8, 9]. Bogan e Thorn [6] sugeriram a teoria de que os HS estariam rodeados por resíduos não importantes para a ligação, cujo papel seria proteger os HS da ação do solvente. Mais tarde, esta teoria ficou conhecida como a “teoria do O-ring”. Esta teoria foi formulada com base em algumas observações efetuadas: resíduos que têm um grande impacto na energia livre de ligação estão fortemente protegidos do contacto com o solvente uma vez que nas estruturas raio-X têm um valor baixo, ou zero, de área acessível ao solvente (accessible solvent area, ASA), e não existem resíduos com grande ASA que contribuam consideravelmente para a $\Delta\Delta G_{\text{ligação}}$. No entanto, existem resíduos ocluídos do solvente que não contribuem

significativamente para a ligação. A oclusão do solvente favorece, desta forma, resíduos (triptofano, tirosina, asparagina) que são capazes de estabelecer ligações de hidrogénio e interações hidrofóbicas. Atualmente, a teoria do O-ring é uma hipótese comumente aceite pela comunidade científica.

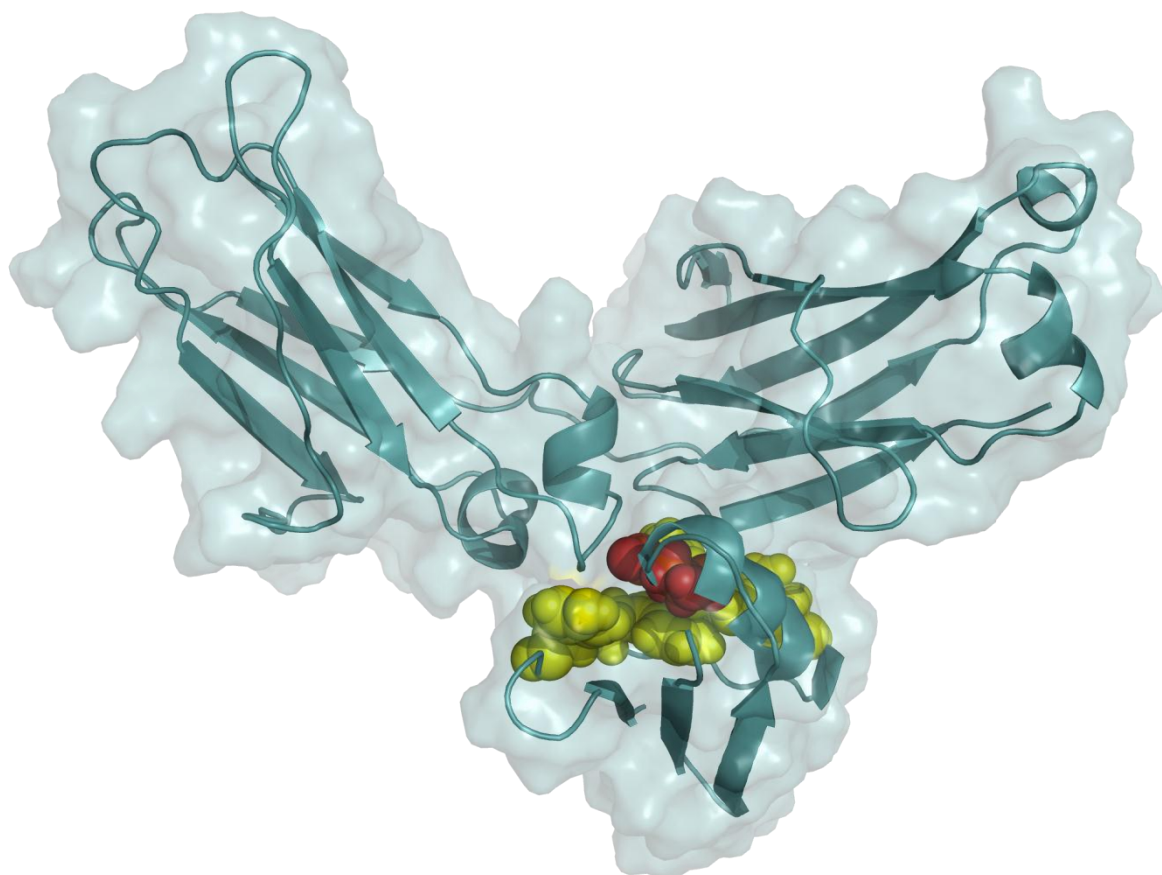


Figura I.1 - Representação de HS (vermelho) e NS (amarelo) numa interface proteína-proteína (PdbID: 1FCC [10]).

O método computacional desenvolvido por *Moreira et al.* para o estudo de HS em interfaces proteína-proteína atingiu uma taxa de sucesso apreciável na determinação de energias livres de ligação, sendo objeto de várias publicações científicas internacionais [2, 7, 9, 11].

1.2. Sistemas proteína-DNA

O estudo da interface e dos aminoácidos que governam a formação de complexos proteína-proteína é longo e decorre há vários anos. Complexos proteína-DNA têm um papel central em todos os aspetos da atividade genética num organismo, como é o caso da transcrição, replicação, recombinação ou reparação. É, por isso, de grande importância examinar a natureza dos complexos formados entre proteínas e DNA, pois eles são a base do nosso

Capítulo I: Introdução

conhecimento de como estes processos ocorrem. Atualmente, existem centenas de complexos proteína-DNA obtidos por cristalografia de raio-X, cujas estruturas providenciam informação dos princípios que governam a ligação, reorganização das sequências de bases e como a estrutura de DNA é modificada após ligação. Estas estruturas experimentais estão disponíveis no Protein Data Bank e em bases de dados específicas para complexos proteína-DNA, como são o caso: AANT, que contém informação estatística de interações aminoácido-nucleótido; ProNIT, que contém informação experimental da ligação de complexos proteína-DNA; PDIdb, que coloca especial importância na interface, entre outras. Com o aumento do número de estruturas cristalográficas de complexos proteína-DNA, tornou-se possível classificar estes complexos com base na sua Classe e no seu Tipo. A categoria **Classe** é baseada na função e contém três subcategorias: *Enzima*, se a função principal da proteína é modificar DNA; *Fator de Transcrição*, se a função principal da proteína é regular a transcrição e expressão do gene; *Proteína de Suporte*, se a função principal da proteína é servir de suporte ao DNA. A categoria **Tipo** é baseada na função / estrutura, e tem 15 subcategorias para a Classe Enzima (Polimerase, Nuclease, Endonuclease, Ligase, Metiltransferase, etc.) 7 subcategorias para a Classe Fator de Transcrição (*Alpha Helix*, *Alpha / Beta*, *Beta Sheet*, *Helix turn Helix*, *Ribbon / Helix / Helix*, *Zinc Coordinating*, *Zipper Type*) e 8 subcategorias para a Classe Proteína de Suporte [12, 13].

A identificação de sistemas proteína-DNA e dos seus locais de ligação pode ser realizada em laboratório (*in vivo* ou *in vitro*) ou utilizando técnicas computacionais (*in silico*). As proteínas interagem com o DNA através de interações eletrostáticas (pontes salinas), interações dipolares (ligações de hidrogénio), interações hidrofóbicas e forças de dispersão. Diferentes tipos de técnicas laboratoriais têm sido desenvolvidas para estudar este tipo de interações complexas entre proteínas e ácidos nucleicos. Entre estas incluem-se ensaios de Imunoprecipitação por Cromatina (*Chromatin Immunoprecipitation*, ChIP), em que as células vivas são tratadas com formaldeído para estabilizar as interações proteína-DNA, permitindo a purificação e deteção; ensaios da mobilidade eletroforética do DNA (*DNA Electrophoretic Mobility Shift Assay*), que podem ser usados para avaliar o grau de afinidade ou especificidade da interação; ensaios de captura e deteção em microplaca (*Microplate Capture and Detection Assay*), entre outros [14].

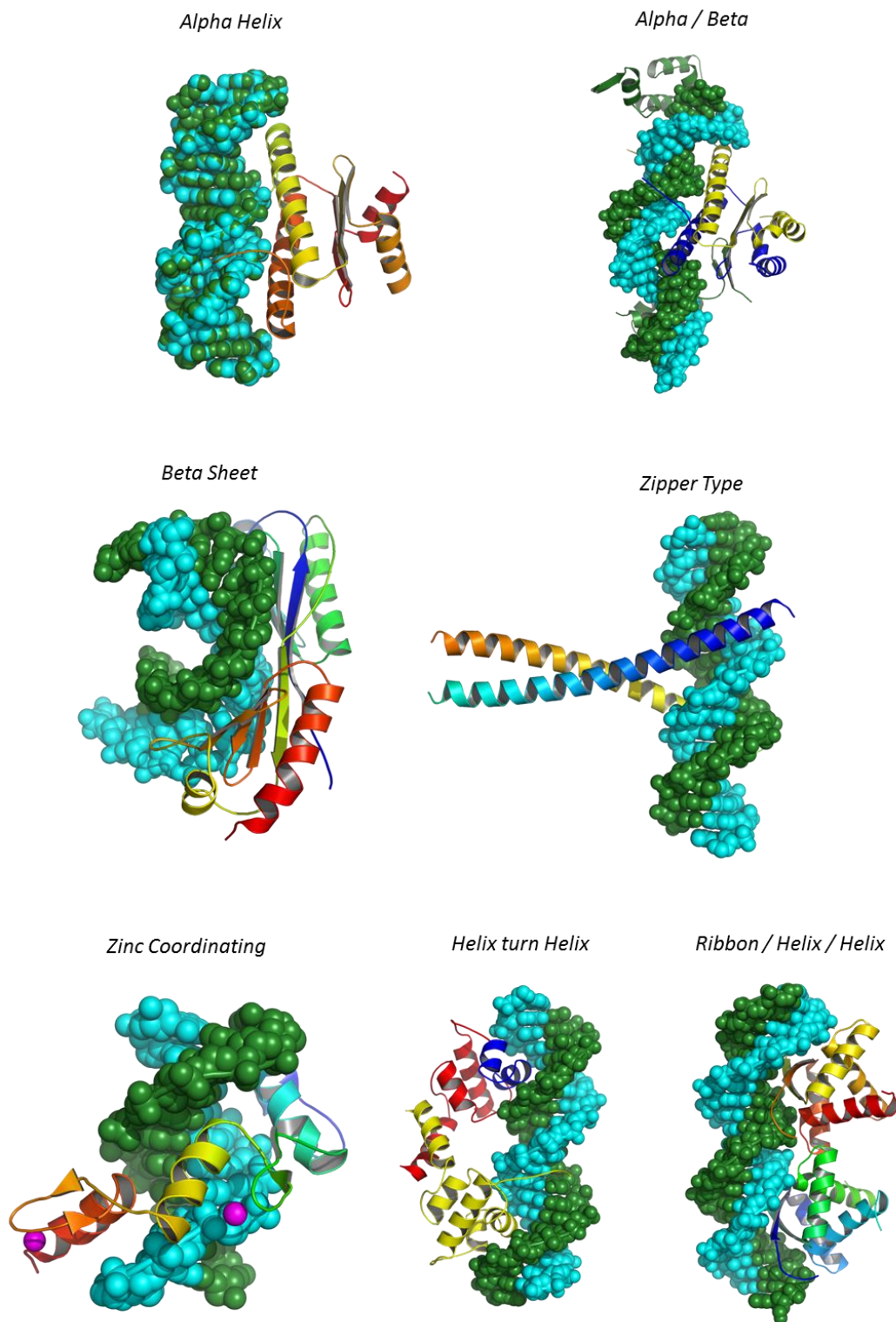


Figura I.2 - Representação dos vários tipos de Fatores de Transcrição. Proteína está representada em cartoon e o DNA em vdW.

Embora já tenham sido desenvolvidas metodologias computacionais capazes de determinar os HS presentes em complexos proteína-proteína, tais não são aplicáveis no caso de interfaces proteína-DNA [2, 11]. A aplicação do método computacional ASM para interfaces proteína-DNA é ainda, do nosso conhecimento, bastante escassa. A identificação de HS em interfaces

Capítulo I: Introdução

proteína-DNA é uma temática ainda pouco desenvolvida, em grande parte devido às dificuldades associadas à simulação de ácidos nucleicos e à própria adaptação do método.

Estudos realizados com base em dados termodinâmicos de mutações em complexos proteína-DNA apontam para a tendência dos resíduos mais conservados se agruparem em *clusters*, em vez de estarem distribuídos ao longo da interface. Esta situação é coincidente com o comportamento verificado em complexos proteína-proteína. Relativamente à distribuição destes resíduos na superfície verificou-se que os resíduos conservados ocorriam com maior frequência para valores baixos de ASA, estando rodeados por outros resíduos que formam fortes redes de interações. Os HS quando agrupados em *clusters* contribuem mais para a estabilidade do complexo proteína-DNA do que quando se encontram isolados, ou seja, existe cooperatividade [15].

Anteriormente referiu-se que a distribuição dos HS num sistema proteína-proteína não era homogênea, mas que estes se organizavam em regiões altamente condensadas. O mesmo tipo de comportamento e organização é encontrado em sistemas proteína-DNA, ou seja, existe um agrupamento dos HS em regiões centrais e não nas extremidades da interface. Similarmente, e apesar das diferenças no carácter geral das interfaces, quando se compara os tipos de aminoácidos existentes nas regiões formadas por HS, verifica-se a existência do mesmo tipo de resíduos nomeadamente, Triptofano, Fenilalanina e Tirosina. Contudo, a elevada ocorrência de resíduos carregados positivamente (Arginina e Lisina) bem como a menor ocorrência de resíduos hidrofóbicos e carregados negativamente, distingue a interface proteína-DNA de uma interface proteína-proteína, respetivamente.

No presente trabalho, procurou-se adaptar o método desenvolvido por *Moreira et al.* [11] para a identificação de HS em interfaces proteína-proteína, para aplicação em complexos proteína-DNA.

2. Um caso particular de um sistema proteína-DNA: HIV-1 RT

2.1. HIV

O vírus da imunodeficiência humana (VIH), ou HIV (sigla em inglês para *human immunodeficiency virus*) é o lentivírus (um membro da família dos retrovírus) responsável pela síndrome da imunodeficiência adquirida (SIDA). Esta síndrome diminui progressivamente a eficácia do sistema imunológico deixando-o desprotegido contra infecções oportunistas. Desde a sua descoberta, em 1981, estima-se que a SIDA tenha levado à morte mais de 46 milhões de pessoas, tornando-a uma das doenças mais mortíferas da história. Atualmente, o número de pessoas infectadas pelo vírus do HIV, tal como o número de mortes provocadas por doenças relacionadas com o vírus, continua a diminuir globalmente, em particular nos países africanos, região mais afetada pela epidemia [16]. O decréscimo da mortalidade causada pelo vírus encontra a sua explicação nas terapias antirretrovirais aplicadas, juntamente com o desenvolvimento de fármacos com a capacidade de bloquear a replicação do vírus.

Foi precisamente num destes sistemas proteína-DNA, formado pela enzima Transcriptase Reversa (RT, sigla em inglês para Reverse Transcriptase), por uma dupla cadeia de DNA e pelo inibidor Tenofovir Disoproxil Fumarate (TDF), que parte do trabalho aqui discutido se centrou.

2.1.1. Transcriptase Reversa e o vírus HIV-1

Na ausência de uma vacina com capacidade para eliminar definitivamente o vírus, a utilização de fármacos surge como a única alternativa terapêutica capaz de tratar este tipo de infeções. Contudo, estas não podem ser eliminadas, por isso, uma vez começado o tratamento, este deve ser mantido ao longo da vida do indivíduo, colocando diversos problemas no *design* de fármacos anti-HIV, uma vez que estes têm de ser relativamente não-tóxicos de forma a serem usados em tratamentos de longa duração. O processo de replicação do HIV-1 é propenso a erros, dando origem a mutações que, combinadas com a rápida replicação do vírus nos indivíduos infectados, causam resistência a todos os fármacos disponíveis para o seu tratamento[17]. Assim, a única forma de evitar o desenvolvimento de resistência é bloquear completamente a replicação viral. Para tal, os tratamentos e os fármacos aprovados atualmente incidem sobre três enzimas, que se sabe serem responsáveis pela replicação do vírus: Transcriptase Reversa (RT), Protease (PR) e Integrase (IN).

Capítulo I: Introdução

A transcrição reversa da cadeia simples de RNA viral para uma dupla cadeia de DNA (dsDNA) é um passo essencial no processo de replicação do HIV. Um conjunto de proteínas e enzimas virais participam no processo de conversão de RNA para DNA; contudo, a transcrição reversa é totalmente dependente da ação da enzima RT[18]. A enzima HIV-1 RT tem, desta forma, a função de converter a cadeia simples de RNA viral numa dupla cadeia de DNA, processo que ocorre no citoplasma da célula infetada. dsDNA é de seguida transportado para o núcleo onde será integrado no genoma do portador pela enzima IN.

2.1.1.1 Constituição e domínios da HIV-1 RT

A HIV-1 RT é um heterodímero assimétrico composto por duas subunidades: p66 e p51. Estas duas subunidades são compostas por 560 e 440 aminoácidos, respetivamente. A HIV-1 RT apresenta dois domínios catalíticos: DNA polimerase e Ribonuclease H (RNase H). A cadeia simples de RNA viral liga-se à polimerase que efetua a transcrição para uma dupla cadeia RNA/DNA; de seguida a RNase H separa a cadeia de RNA da cadeia de DNA, voltando esta última à polimerase para ser formada a dupla cadeia de DNA. Estes dois domínios, cooperando entre si, são os responsáveis pela conversão da cadeia simples de RNA em dupla cadeia de DNA. O domínio da polimerase é composto por quatro subdomínios: *fingers* (resíduos 1-85, 118-155), *palm* (resíduos 86-117, 156-236), *thumb* (resíduos 237-318) e *connection* (resíduos 319-426) [19-21]. Os domínios da p66 foram assim designados pois a sua estrutura é semelhante à de uma mão. A estrutura da HIV-1 RT, com os seus domínios e subdomínios, encontra-se representada na Figura I.3.

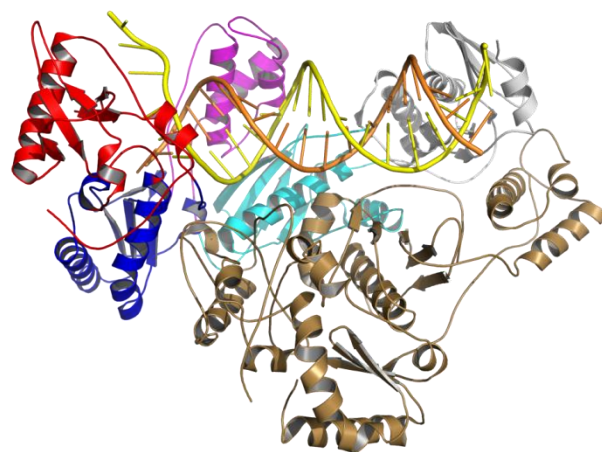


Figura I.3 - Representação da HIV-1 RT (PDBID:1T05). *Fingers*, *palm*, *thumb*, *connection* e a RNase H da subunidade p66 estão representados a azul, vermelho, rosa, ciano e cinzento, respetivamente. A subunidade p51 encontra-se representada a castanho. O template e o primer de DNA estão representados a amarelo e laranja.

2.1.1.2 Mecanismo de inibição / resistência da HIV-1 RT

Devido ao papel que desempenha no ciclo de reprodução do vírus HIV-1, a enzima RT tem sido um alvo fundamental na criação de terapias anti-HIV. Por esta razão, grande parte dos fármacos utilizados no combate à SIDA procuram inibir a atividade da polimerase da RT. Inibidores específicos para a RNase H são menos comuns e, apesar de alguns já se encontrarem descritos, ainda não foram aprovados para o tratamento de infecções [22-24].

O aparecimento de novas estirpes do vírus, resistentes aos fármacos, é uma consequência da alta replicação, da incompleta supressão do vírus e de uma síntese de DNA muito suscetível a erros. Estas situações acabam por originar mutações que tornam a RT resistente aos inibidores. Até ao momento, conhecem-se duas estratégias usadas pela RT de forma a desenvolver resistência aos inibidores: (i) Interferência na incorporação dos inibidores – neste mecanismo de resistência mutações em resíduos do centro ativo, sobretudo nos subdomínios *finger* e *palm* da p66, impedem a correta incorporação do inibidor; (ii) Excisão – ocorre quando a RT incorpora eficientemente o inibidor, que bloqueia temporariamente a síntese de DNA, mas é posteriormente removido (excisado) originando uma 3'-OH livre que permite o retomar da síntese. A reação de excisão requer um dador de pirofosfato (PPi) de forma a poder libertar o inibidor do primer. Apesar de tanto o ATP como o PPi poderem atuar como dadores de pirofosfato, sabe-se que é o ataque nucleófilo do ATP à extremidade do primer que remove o inibidor e origina uma 3'-OH livre. Quando uma dupla cadeia template / primer se liga à RT com a 3' final do primer situada no P-site, o N-site está disponível para se ligar ao próximo dNTP. Desta forma, após a ligação do próximo dNTP, forma-se um complexo ternário terminal, muito estável, com o dNTP no N-site e a extremidade do primer no P-site. Como a extremidade do primer se encontra no P-site, a excisão não pode ocorrer. Por outro lado, se esta estivesse localizada no N-site no chamado complexo “pré-translacional”, poderia ocorrer o ataque nucleófilo por parte do ATP que consumava a excisão.

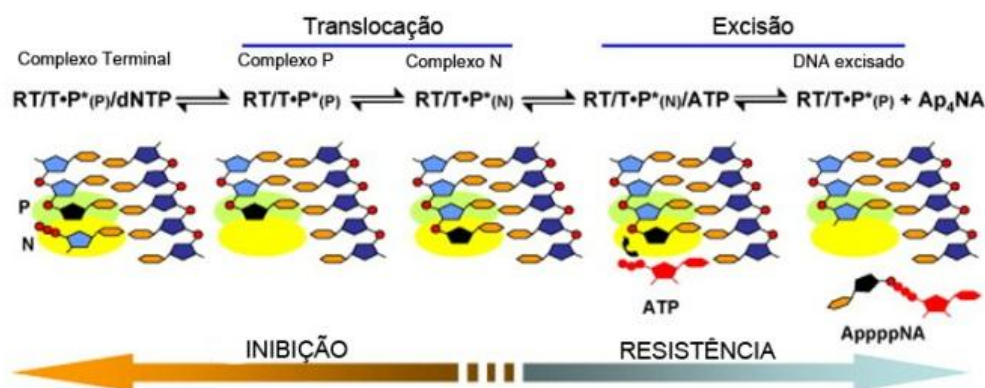


Figura I.4 - Modelo simples representativo do equilíbrio entre o processo de translocação e de excisão. A inibição da síntese de DNA, devido à formação de um complexo terminal, é favorecida quando o primer se encontra ligado no P-sítio. Adaptado (Menéndez-Arias, L. [25]).

2.1.2. Tipos de Inibidores da HIV-1 RT

Diferentes tipos de inibidores da HIV-1 RT estão disponíveis atualmente, tendo a sua utilização sido aprovada para o combate à infecção. Muitos outros encontram-se nas diversas fases de testes clínicos. Os inibidores da HIV-1 RT podem ser agrupados em dois grandes grupos: NRTIs (do inglês, Nucleoside Reverse Transcriptase Inhibitors) e NNRTIs (do inglês, Non-nucleoside Reverse Transcriptase Inhibitors). NRTIs são a classe principal de inibidores da HIV-1 RT, e deste grupo fazem parte inibidores como: AZT (zidovudine, Retrovir[®]), d4T (stavudine, Zerit[®]) ou ddI (didanosine, Videx[®]). Estes inibidores, depois de entrarem na célula, precisam de ser fosforilados para a sua forma trifosfática (TP) de forma a atuarem como inibidores competitivos do substrato natural, dNTP[26]. Depois de incorporados pela RT, atuam como terminadores de cadeia, uma vez que lhes falta um grupo 3'-OH, não permitindo a adição de mais nucleótidos à cadeia. Atualmente, todos os NRTIs atuam como terminadores de cadeia após a sua incorporação no DNA viral pela RT. NNRTIs ligam-se a um centro hidrofóbico da HIV-1 RT, próximo do centro ativo da polimerase, na base da subunidade *thumb*. A comparação de estruturas na presença e na ausência do inibidor mostram que este centro não existe na ausência do inibidor [27]. É criado com a ligação do inibidor devido a uma alteração conformacional da cadeia lateral de resíduos da RT. Ao contrário dos NRTIs, os NNRTIs são inibidores não-competitivos, relativamente ao substrato natural (dNTP), baseando a sua ação na distorção de resíduos do centro ativo da polimerase e da subunidade *thumb* da HIV-1 RT, que impedem o correto posicionamento do primer. O deslocamento dos resíduos do centro ativo acaba por afetar a síntese de DNA [19, 28]. Esta

classe inclui inibidores tais como o nevirapine (Viramune[®]), delavirdine (Rescriptor[®]) ou efavirenz (Sustiva[®], Stocrin[®]). É importante fazer uma distinção entre NRTIs e NtRTIs (do inglês, Nucleotide Reverse Transcriptase Inhibitors). NtRTIs, apesar de similares e de serem muitas vezes englobados e considerados juntamente com os NRTIs, apresentam algumas diferenças, quer na sua estrutura, quer no seu processo de fosforilação. Nos NRTIs, o grupo fosfato é ligado ao nucleósido por uma ligação éster, algo que acontece durante o primeiro passo de fosforilação devido à ação de quinases enzimáticas. No caso dos NtRTIs, o grupo fosfato, na forma de um grupo fosfonato, já se encontra ligado ao nucleósido, o que resulta na formação de um éter fosfometilado, mais estável ao possível ataque de enzimas catabólicas [29, 30]. Os NtRTIs atuam como terminadores de cadeia, mas apenas precisam de duas etapas de fosforilação para atingirem a sua forma ativa (difosfática, DP) uma vez que já contêm um grupo fosfato na sua estrutura. Esta acaba por ser uma vantagem importante, pois permite-lhes evitar a primeira etapa de fosforilação, que é a etapa mais crítica do processo. O principal inibidor desta classe é o TDF (Tenofovir, Viread[®]).

A ação dos inibidores é afetada por três fatores principais: (i) a eficiência pela qual eles são convertidos na sua forma ativa, que no caso dos NRTIs é a trifosfática. O TDF, como já contém um grupo fosfato, apenas necessita da adição de dois fosfatos; (ii) necessitam de ser razoavelmente estáveis a enzimas catabólicas existentes na célula hospedeira; (iii) a sua forma ativa tem de inibir eficientemente a HIV-1 RT.

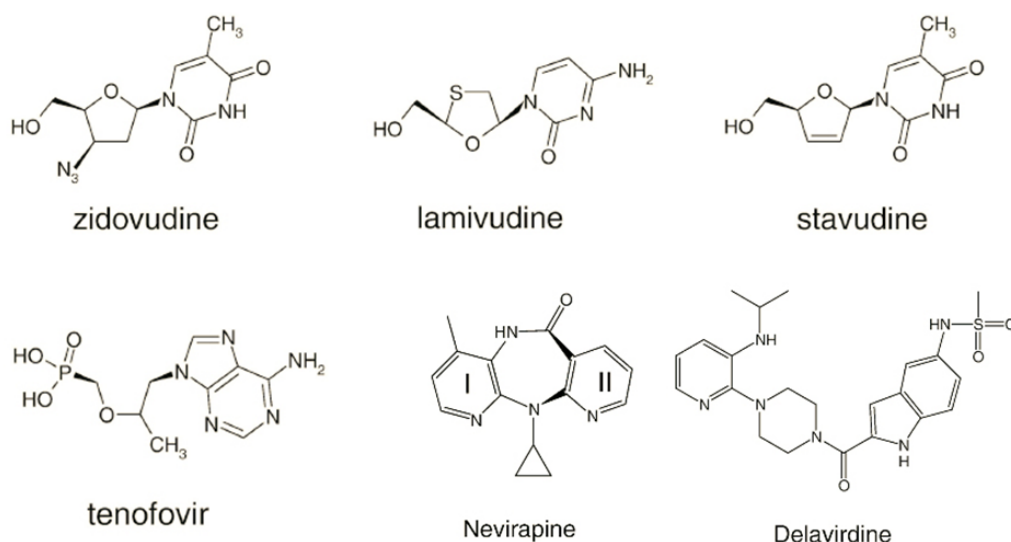


Figura I.5 - Estrutura química de vários NRTIs e NNRTIs.

Capítulo I: Introdução

2.1.3. Tenofovir

Tenofovir Disoproxil Fumarate (TDF), comercializado com o nome Viread[®], da classe dos NtRTIs, é um dos principais inibidores da enzima RT. Foi descrito pela primeira vez em 1993[31] e foi introduzido na prática clínica em 2001, após aprovação por parte da *U.S. Food and Drug Administration (FDA)* para o tratamento do vírus HIV. Mais tarde foi, também, aprovado para o tratamento da hepatite B.

O TDF, tal como foi referido anteriormente, pertence ao grupo dos NtRTIs, uma vez que apresenta diferenças estruturais e funcionais relativamente aos NRTIs. Como pode ser observado na Figura I.5, que apresenta a estrutura de vários inibidores da HIV-1 RT, o TDF demarca-se dos restantes por ter um *linker* acíclico em vez do anel aromático presente nos restantes NRTIs. Outra grande diferença é o facto do TDF já ter incluído na sua estrutura um grupo fosfato. Com base na sua estrutura, pode-se dividir o TDF em três zonas fundamentais: (i) a base, que no caso do TDF é uma adenina; (ii) o linker acíclico e (iii) o grupo fosfonato. Como foi referido anteriormente, o facto de o TDF conter na sua estrutura um grupo fosfato, na forma de um grupo fosfonato, faz com que este necessite apenas de duas etapas de fosforilação para ficar na sua forma ativa (Tenofovir di-fosfato, TDF-DP). Além da vantagem de evitar a primeira etapa de fosforilação, existe um aumento de estabilidade induzida pela presença do grupo fosfonato que torna mais difícil a ação das enzimas catabólicas, depois da ligação à 3'-terminal da cadeia de DNA, quando comparado com os NRTIs e dNTPs.

A sua função é a mesma de qualquer outro NRTI: bloquear a síntese de DNA efetuada pela HIV-1 RT. Depois da sua administração e entrada na célula, é convertido em TDF-DP por quinases enzimáticas, atuando como inibidor competitivo dos dNTPs, e após incorporação termina a síntese de DNA viral devido à formação de um complexo terminal.

2.1.4. Mutações que conferem resistência à HIV-1 RT

O estudo e desenvolvimento de fármacos com capacidade para bloquear a síntese de DNA contendo o vírus HIV-1 já ocorre há mais de duas décadas. Contudo, ainda não foi possível desenvolver um inibidor que eliminasse completamente o vírus. Em grande parte, esta situação deve-se à constante replicação do vírus e ao facto do processo de síntese ser falível e originar mutações. Estas mutações tornam a HIV-1 RT mais resistente à ação dos inibidores, seja por afetarem o mecanismo de incorporação ou o de excisão, os dois mecanismos

preferenciais de resistência da HIV-1 RT, tal como foi anteriormente discutido. Resíduos que estão situados na vizinhança do centro ativo da polimerase são aqueles cuja mutação pode conferir maior resistência à RT. Nestes resíduos incluem-se, por exemplo, a Lys65, Arg72, Asp113, Ala114, Tyr115 ou Gln151 que estão envolvidos diretamente na interação com o ligando, enquanto outros como a Leu74, Pro157, Phe160, Tyr183 ou Met184 podem afetar indiretamente a ligação do ligando (seja ele um inibidor ou um dNTP) à RT [19]. O resíduo Lys65 está localizado no subdomínio *fingers* da RT, podendo um dos seus grupos amino interagir com o último grupo fosfato, quer de um dNTP ou com o TDF-DP [19, 32]. A mutação K65R é a principal mutação causadora de resistência ao TDF [33], sendo também ela responsável por resistência aos outros NRTIs com a exceção do AZT [34]. A mutação K65R reduz a taxa de incorporação (k_{pol}) de dNTPs e de NRTIs e reduz também a capacidade de replicação do vírus, sendo que a taxa de incorporação para os NRTIs é inferior à dos dNTPs. Após a cristalização da estrutura da K65R RT com o TDF-DP ou dATP, foi levantada a hipótese da menor percentagem de incorporação do TDF se dever a uma restrição conformacional dos resíduos Arg65 e Arg72. A mutação causa também uma diminuição na percentagem de excisão.

3. Metodologia Computacional

A química e a bioquímica computacional são áreas científicas que têm tido um crescimento muito acentuado nas últimas décadas, devido à sua capacidade para estudar uma grande variedade de processos biológicos. Devido a uma simbiose quase perfeita entre *hardware* e *software*, a química computacional permite-nos simular uma variedade enorme de sistemas químicos e bioquímicos, bem como descrever propriedades específicas de um dado sistema, tais como a estrutura molecular, as propriedades eletrónicas dos átomos, a relação entre átomos e moléculas, as interações entre moléculas, a geometria molecular (distâncias, ângulos, etc.) ou aferir sobre os estados de transição numa reação. Com a evolução do conhecimento nesta área, os resultados destes cálculos são muitas vezes de maior confiança do que os próprios resultados experimentais. O facto da química computacional nos permitir olhar para um sistema sob diferentes níveis teóricos, sejam eles a quântica ou a mecânica molecular, bem como a possibilidade de estudar dinamicamente um sistema (deixamos de estar condicionados a estruturas rígidas) através da Dinâmica Molecular, tornam-na um aliado fundamental para qualquer químico ou bioquímico.

Dentro da química computacional, uma área tem vindo a destacar-se nos últimos anos: a bioinformática. E o que é a bioinformática? De seguida, transcreve-se a definição de bioinformática submetida no dicionário inglês de Oxford (Oxford English Dictionary).

“ (*Molecular*) **bio – informatics**: bioinformatics is conceptualizing biology in terms of molecules (in the sense of Physical chemistry) and applying “**informatics techniques**” (derived from disciplines such as applied maths, computer science and statistics) to **understand** and **organize** the **information** associated with these molecules, on a **large scale**. In short, bioinformatics is a management information system for molecular biology and has many **practical applications**. “

A bioinformática tem demonstrado ser uma poderosa aliada na compreensão e interpretação de fenómenos biológicos como as interações proteína-proteína [2], proteína-ligando [35], bem como na descoberta de novos fármacos. O processo de descoberta e lançamento para o mercado de novos fármacos envolve um grande número de etapas e vem associado a um custo e tempo necessário muito elevados. O uso de técnicas computacionais de forma a identificar novos fármacos com a atividade pretendida (*hit / lead*), bem como a otimização dos mesmos, pode ajudar na descoberta de candidatos e na colocação no mercado de forma mais rápida e com menos custos associados [36, 37].

As metodologias computacionais mais utilizadas podem ser divididas em quatro grandes grupos: (i) métodos baseados na função de onda – estes métodos podem ser subdivididos em *ab initio* e *semi-empíricos*, sendo aqueles que proporcionam os resultados mais precisos e rigorosos, pois são aqueles que oferecem a melhor aproximação matemática ao sistema estudado. São, contudo, os métodos mais exigentes do ponto de vista computacional e só podem ser aplicados a sistemas de tamanho não muito elevado. Procuram resolver a equação de Schrödinger utilizando a mecânica quântica; (ii) métodos baseados na densidade eletrônica (*Density Functional Theory, DFT*) – tal como o método anterior, baseia-se na mecânica quântica e, tal como este, é muito exigente do ponto de vista computacional. Procuram determinar as propriedades de sistemas polieletrônicos usando funcionais, isto é, funções de outra função, que neste caso é a densidade eletrônica; (iii) métodos baseados na mecânica clássica e (iv) métodos híbridos QM/MM – estes métodos, como o seu nome indica (*Quantum Mechanics / Molecular Mechanics*), são métodos híbridos que permitem combinar as vantagens da mecânica quântica (precisão) e da mecânica molecular (velocidade) no estudo de processos químicos em solução ou em proteínas. É comum tratar-se a zona mais pequena do sistema, como por exemplo o núcleo, por mecânica clássica e o resto do sistema por mecânica molecular. Neste trabalho, utilizaram-se, fundamentalmente, métodos baseados em mecânica clássica que irão ser descritos ao longo desta introdução.

3.1. Mecânica Molecular e Campos de Força

Fator fundamental para o sucesso de qualquer metodologia computacional no estudo de sistemas químicos é a qualidade do modelo matemático usado para calcular a energia do sistema como função da sua estrutura. Para sistemas químicos pequenos, estudados na fase gasosa, métodos baseados na mecânica quântica são apropriados. O tempo de computação associado a estes métodos é muito elevado e dependente do número de eletrões e átomos, e por isso, a sua utilização está geralmente limitada a sistemas contendo aproximadamente 150 átomos. Estes métodos não seriam apropriados para o estudo de sistemas bioquímicos constituídos por um grande número de macromoléculas, como as proteínas, e nos quais se utiliza habitualmente um solvente explícito, resultando num tamanho final do sistema na ordem dos milhares de átomos. Por esta razão, é comum utilizarem-se métodos de Mecânica Molecular (MM) em bioquímica computacional. Os recursos necessários, a nível de poder de computação, são menores o que lhes permite estudar sistemas com uma ordem de magnitude que não seria alcançada pelos métodos quânticos. A MM faz uso das leis da mecânica

Capítulo I: Introdução

clássica, através das leis de Newton, para descrever o movimento das partículas. De uma forma simplista, uma molécula pode ser considerada como certo número de pesos ligados entre si por molas, sendo que os pesos representam os núcleos e as molas representam as ligações. É sabido que a massa do núcleo é bastante superior à massa do elétron, e que a sua velocidade é muito pequena quando comparada com a velocidade dos elétrons (aproximação de Born-Oppenheimer). Assim, a aproximação que é feita é que os elétrons se adequam quase imediatamente aos movimentos do núcleo. Esta aproximação faz com que se possam considerar apenas os movimentos nucleares como os importantes para descrever a molécula. Uma vez que o método não considera os elétrons, não consegue definir um conjunto de propriedades químicas muito importantes, como são o caso da quebra ou formação de ligações e estados eletrônicos excitados. Esta é, provavelmente, a maior desvantagem do método que não consegue assim estudar a evolução de reações químicas (daí nestes casos se utilizarem os métodos híbridos QM/MM).

Na MM o átomo é a partícula mais pequena do sistema, sendo os núcleos e elétrons tratados através da parametrização de Campos de Força (CF). Um CF é definido, no contexto da MM, como sendo uma expressão matemática de variáveis físicas e o conjunto de parâmetros usados para descrever a energia potencial de um sistema. Como iremos ver mais adiante, existe um grande número de CF disponíveis para tratar sistemas químicos e bioquímicos; contudo, todos eles partilham genericamente a mesma expressão matemática para o cálculo da energia:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + E_{cross} \quad (1)$$

Como pode ser observado na equação 1, a energia total é dada como uma soma de termos que representam uma contribuição física diferente. Diferentes formulações matemáticas podem existir para cada termo, de acordo com o respetivo CF, com diferenças na precisão e no tempo de computação. Os três primeiros termos (E_{str} , E_{bend} e E_{tors}), que estão relacionados com átomos ligados por uma ligação covalente, descrevem a energia para o estiramento na ligação entre dois átomos (*stretching*); a energia necessária para fazer variar um ângulo formado por três átomos (*bending*) e a energia torsional para os ângulos diedros (*torsion*), respetivamente. Os dois termos seguintes (E_{vdw} e E_{elec}), interações não-covalentes, descrevem as interações resultantes de forças de van der Waals e forças eletrostáticas. O último termo (E_{cross}) correlaciona os termos mencionados anteriormente, mas é muitas vezes omitido (e não utilizado) por causar um aumento no tempo de computação. Os fatores que constituem esta expressão de energia juntamente com um conjunto de parâmetros empíricos, geralmente

obtidos de dados experimentais ou de cálculos efetuados para moléculas pequenas a um nível teórico superior (por exemplo HF ou DFT), compõem o CF.

Existem, atualmente, múltiplas variedades de CF disponíveis, entre os quais: MM2/3/4 (Molecular Mechanic Force Field for Small Molecules)[38], CHARMM (Chemistry at Harvard Macromolecular Mechanics)[39, 40], AMBER (Assisted Model Building with Energy Refinement)[41], OPLS (Optimized Parameters for Liquid Simulation)[42, 43], CFF (Consistent Force Field)[44], GROMOS (Groningen Molecular Simulation packages)[45], entre outros. Uma vez que existem tantos CF diferentes, é natural que cada um deles tenha potencialidades diferentes, ou que se adapte melhor a determinado tipo de sistema. Aqueles mais utilizados na descrição de proteínas e ácidos nucleicos são o AMBER e o CHARMM. Neste trabalho, foi utilizado o campo de forças GAFF (General Amber Force Field)[46] em conjunto com o CF do AMBER *ff99SB*. O GAFF foi desenvolvido com o intuito de complementar o AMBER, já conhecido pela sua capacidade de simular sistemas com proteínas e ácidos nucleicos, adicionando parâmetros para praticamente todas as moléculas orgânicas. Este CF, compatível com o AMBER, é assim um forte aliado no *design* de novos fármacos, pois funciona para a maioria das moléculas farmacêuticas.

$$\begin{aligned} & \sum_{stretching} K_r (r - r_{eq})^2 + \sum_{bending} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsional} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \\ & \sum_{non-bonded} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \end{aligned} \quad (2)$$

Analisando a equação 2, equação típica para o GAFF, e à semelhança com o descrito na equação 1, temos termos que representam interações covalentes e termos que representam interações não-covalentes. Será analisado de seguida, de forma mais pormenorizada, cada um destes termos.

O primeiro termo da equação 2, referente ao estiramento da ligação entre dois átomos (*stretching*), está relacionado com a variação de energia que acontece aquando do aumento ou contração de ligações entre as moléculas. A grande maioria dos CF descreve esta relação segundo um potencial harmónico, por aplicação da Lei de Hooke (equação 3).

$$V_{stretching} = \frac{1}{2} K_r (r - r_{eq})^2 \quad (3)$$

Capítulo I: Introdução

O oscilador harmónico é um modelo mecânico simples de uma massa em movimento, que se encontra fixa a uma parede por meio de uma mola (Figura I.6). A Lei de Hooke dá-nos a relação entre a força aplicada a uma mola não esticada e o deslocamento da mola depois da aplicação de uma força. Assim sendo, a elongação ou

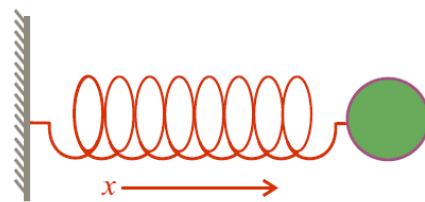


Figura I.6 - Ilustração do oscilador harmónico.

compressão de uma ligação é caracterizada por uma constante de força, K_r , multiplicada pelo quadrado da elongação da ligação, em que r representa o comprimento da ligação quando esta é deformada e r_{eq} representa a distância da ligação quando esta se encontra próxima do valor de referência, sendo uma aproximação ao valor de equilíbrio da ligação. Outra forma de descrever o estiramento da ligação entre dois átomos seria por aplicação do potencial de Morse (equação 4).

$$V_l = D_e \{1 - \exp[-a(l - l_0)]\}^2 \quad (4)$$

Na equação 4, D_e é o potencial de energia mínima, l é a distância entre os dois átomos, l_0 é a distância referência entre os átomos e a é uma quantidade dada pela razão entre a frequência de vibração da ligação, pela massa reduzida e pelo potencial de energia mínima. Apesar do potencial de Morse ser aquele que melhor representa o comportamento da ligação, é raramente incluído nos CF mais comuns, pois é difícil de tratar computacionalmente e requer três parâmetros para cada ligação (D_e , a e l_0) em comparação com os dois necessários pelo potencial harmónico (K_r e r_{eq}). Assim, e uma vez que o potencial harmónico descreve, de forma razoável, o comportamento da ligação na região de valores perto do ponto de referência (associado ao mínimo de energia), juntamente com a necessidade de apenas se determinarem dois parâmetros (o que diminui o tempo de computação), o potencial harmónico configura a melhor opção para descrever o estiramento da ligação. É importante referir que, para zonas afastadas do valor de referência, o potencial harmónico não descreve o melhor comportamento para a ligação, sendo necessária a introdução de termos de ordem superior (Figura I.7).

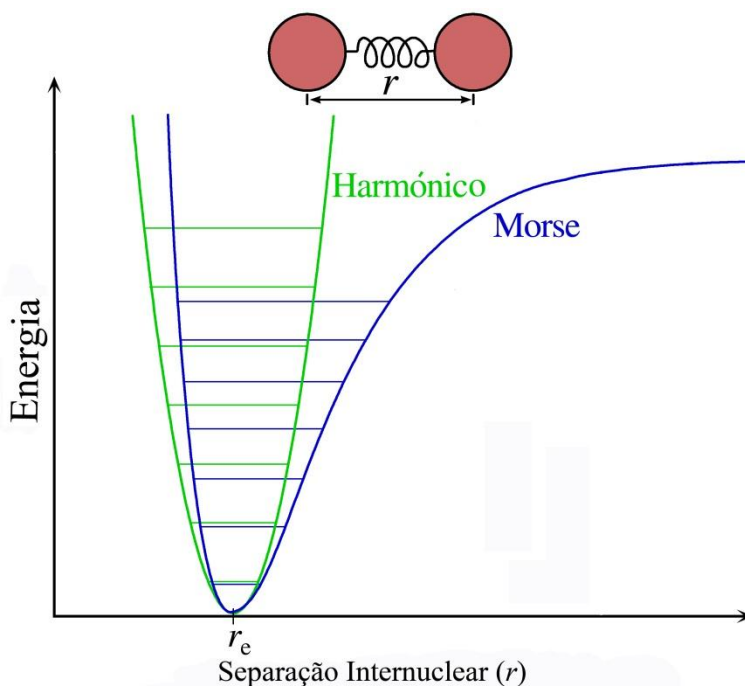


Figura I.7 - Representação gráfica da forma dos potenciais de Morse e Harmônico.

O termo seguinte da equação 2 (*bending*) descreve a variação de energia associada ao aumento ou diminuição de ângulos, formados entre três átomos covalentemente ligados, dos seus valores de referência. Tal como na situação anterior, é normalmente aplicado um potencial harmônico. Na equação seguinte, K_θ é a constante de força associada ao modo de *bending*, θ e θ_{eq} são respetivamente, o ângulo entre os três átomos e o valor de referência (ou de equilíbrio). Tal como no caso anterior, a adição de termos de ordem superior melhora a descrição do comportamento químico em questão, com natural consequência no tempo de computação.

$$V_{bending} = \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 \quad (5)$$

O terceiro termo da equação 2 corresponde à energia resultante da variação do ângulo diedro entre átomos adjacentes na molécula (*torsional*). Uma possível expressão matemática para o cálculo desta energia tem a forma:

$$V_{torsional} = \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)] \quad (6)$$

Capítulo I: Introdução

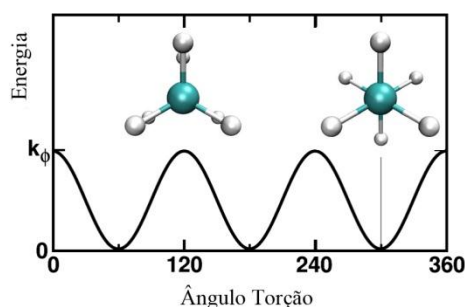


Figura I.8 – Representação da variação da energia torsional para a molécula de etano, como resultado da rotação do ângulo diedro. O máximo corresponde à conformação eclipsada e o mínimo à conformação alternada.

em

que, V_n é a constante de força de torsão; θ é o ângulo diedro, isto é, o ângulo de torção entre a ligação AB e a ligação CD; n é a multiplicidade, ou seja, o número de mínimos da função enquanto a ligação é rodada 360° e γ é o valor mínimo para o ângulo de torção. Esta variação é cíclica tendo máximos quando átomos estão demasiado próximos uns dos outros e mínimos quando estes se afastam (Figura I.8).

Terminados os termos que envolvem ligações covalentes, aparecem de seguida aqueles que definem as interações entre átomos não-ligantes, nomeadamente as interações de van der Waals e as interações eletrostáticas. Começemos por analisar as interações eletrostáticas. Este termo descreve as interações não-ligantes resultantes da presença de cargas atômicas, sendo normalmente descrita pela equação de Coulomb (equação 7).

$$V_{electrostatic} = \sum \frac{q_i q_j}{\epsilon r_{ij}} \quad (7)$$

A energia eletrostática é uma função das cargas atômicas (q_i, q_j), da distância entre estas (r_{ij}) e da constante dielétrica (ϵ). A maior diferença no cálculo do termo da energia eletrostática entre os vários CF disponíveis, reside na forma como as várias cargas atômicas são calculadas. Tal pode ser efetuado utilizando mecânica quântica ou métodos semi-empíricos. De acordo com o método escolhido, diferentes tipos de cargas atômicas estão disponíveis. Exemplos incluem as cargas de Mulliken, as cargas de Merz-Kollman, as cargas de Gasteiger e as cargas RESP (Restrained ElectroStatic Potencial charges) que foram as utilizadas neste trabalho. Este método considera que átomos equivalentes têm a mesma carga pontual.

O termo correspondente à energia de van der Waals descreve as interações (atrações e repulsões) entre átomos que não estejam diretamente coordenados. Estas interações são atrativas a curtas distâncias mas tendem rapidamente para zero quando a distância entre átomos aumenta. Estão incluídas interações intermoleculares entre dois dipolos permanentes,

um dipolo permanente e o correspondente dipolo induzido e entre dois dipolos induzidos. É geralmente utilizado o potencial de Lennard-Jones para calcular este termo (equação 8).

$$V_{vdw} = \sum \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \quad (8)$$

Esta equação representa um potencial de Lennard-Jones 12-6, em que ϵ corresponde à energia mínima da função; A e B são a distância na qual o mínimo da curva de Lennard-Jones ocorre e r é a distância entre os dois átomos. O primeiro e segundo termos da equação de Lennard-Jones são, respetivamente, termos atrativos e repulsivos. Existem outros potenciais que podem ser utilizados para descrever estas interações, como o potencial 10-9; contudo o potencial 12-6 é o mais utilizado. A razão que está por trás desta maior utilização, mesmo havendo melhores potenciais para descrever a interação, tem a ver com facilidade de cálculo uma vez que o resultado do potencial 12 é obtido pelo quadrado do termo atrativo. A Figura I.9 apresenta uma representação típica do potencial de Lennard-Jones.

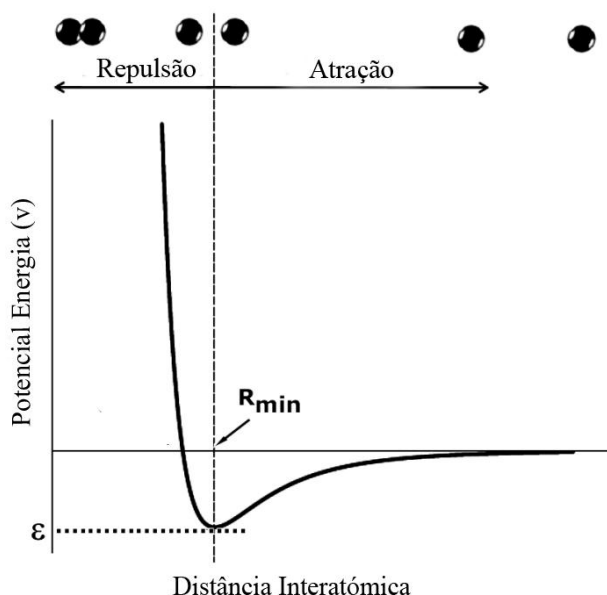


Figura I.9 – Representação gráfica de um potencial de Lennard-Jones 12-6. ϵ é a energia mínima da função e R a distância na qual o mínimo da curva ocorre. Para distâncias interatômicas grandes existe uma tendência para zero e para curtas distâncias repulsões muito acentuadas.

3.2. Minimização da Energia

A minimização de energia é um dos métodos mais simples para otimizar uma dada estrutura. Podem variar em precisão e eficiência computacional, sendo utilizados como uma ferramenta (juntamente com outros métodos computacionais) na preparação de modelos, preparação da

Capítulo I: Introdução

estrutura inicial para dinâmica molecular, entre outros. Os métodos de minimização podem ser divididos em: métodos de ordem zero, métodos de primeira ordem e métodos de segunda ordem. O objectivo de todos os algoritmos de minimização é o de encontrar o mínimo local de uma dada função. Estes algoritmos são classificados de acordo com a quantidade de informação considerada da função usada, representado pela derivada de ordem superior usada pelo algoritmo. Assim sendo, algoritmos que apenas utilizam o valor da função e nenhuma derivada são chamados “algoritmos de ordem 0”; algoritmos que utilizam a primeira derivada são denominados “algoritmos de ordem 1”; finalmente, “algoritmos de ordem 2” são algoritmos que utilizam a segunda derivada. Tal como nas técnicas computacionais, algoritmos de ordem superior que utilizam maior informação da função são geralmente mais precisos, sendo mais exigentes computacionalmente, em comparação com métodos de ordem inferior.

Algoritmos de ordem 1, que representam um balanço razoável entre precisão e eficiência, são os métodos de minimização mais utilizados na simulação de macromoléculas. Fazem uso do declive da função, mas não consideram a curvatura. Para compensar esta deficiência na informação da curvatura, todos os métodos de minimização de ordem 1 utilizam um esquema iterativo passo a passo. Estas iterações são utilizadas para recalcular o gradiente e corrigir o padrão de abordagem da minimização de acordo com alterações na direcção do declive. Dois dos métodos de minimização de ordem 1 mais utilizados são o *Steepest descent* e o *Conjugate Gradient*.

3.3. Dinâmica Molecular

As simulações de Dinâmica Molecular (DM) providenciam uma descrição detalhada de sistemas complexos numa vasta gama de escalas temporais. Esta técnica foi introduzida na década de 50 por dois físicos teóricos, Alder e Wainwright [47]. A primeira publicação de uma simulação de DM para uma proteína aconteceu alguns anos mais tarde, na década de 70, quando *McCammon et al.* estudaram o inibidor pancreático da tripsina bovina (58 resíduos)[48]. Desde essa altura, a DM sofreu uma evolução acentuadíssima e é hoje uma das ferramentas mais utilizadas no estudo de biomoléculas, sendo frequentemente utilizada para investigar a estrutura, flexibilidade e uma série de propriedades de complexos biológicos. No presente, para além de proteínas, uma vasta gama de diferentes biomoléculas é estudada com recurso a simulações de DM, entre as quais os carboidratos, ácidos nucleicos ou sistemas lipídicos. Fator preponderante nesta evolução foi o desenvolvimento computacional

verificado, que permitiu ultrapassar limitações na dimensão do sistema estudado e no tempo de simulação. Atualmente, já é possível realizar simulações de DM em sistemas constituídos por átomos na ordem da centena de milhar. Além disso, simulações que anteriormente demoravam um mês a ser concluídas, são hoje realizadas em menos de um dia. Os sistemas biológicos não são estáticos, bem pelo contrário, apresentam uma grande variedade de movimentos que são importantes para o sistema e que devem ser contabilizados quando o estamos a estudar. Alguns destes incluem movimentos locais, como as flutuações atômicas e o movimento das cadeias laterais, ou movimentos globais como a associação / dissociação de subunidades. A simulação de DM permite analisar os movimentos individuais e coletivos dos constituintes do sistema ao longo do tempo, fornecendo-nos dados que nos permitem interpretar de forma mais clara os vários processos que acontecem. Tal como em todas as áreas da Ciência, é recomendável que haja uma comunhão com a parte experimental de forma a melhor se avaliarem e interpretarem os dados[49].

Neste trabalho, utilizaram-se simulações de DM clássica, que, tal como o próprio nome indica, é baseada nos princípios da Mecânica Molecular e como tal nas Leis de Movimento de Newton. Para além deste tipo de DM, existem ainda mais dois: DM quântica e DM híbrida (QM/MM). As simulações de DM quântica são mais exigentes a nível computacional e como tal são apenas utilizadas para simular sistemas em que efeitos eletrónicos são relevantes, como, por exemplo, a quebra e a formação de ligações.

3.3.1. Ensembles e Equações do Movimento

Simulações computacionais geram informações ao nível microscópico e a conversão desta informação em termos macroscópicos é feita pela utilização da termodinâmica estatística. As simulações de DM geram um conjunto de configurações sucessivas de cada partícula do sistema de forma a obter um conjunto de microestados cronologicamente ordenados. Este conjunto de microestados cronologicamente ordenados define um macroestado (estado termodinâmico de um sistema) que pode ser caracterizado pela especificação de funções de estado (como o volume, temperatura, pressão, etc.). Daqui surge a noção de ensemble, que corresponde a um conjunto de microestados que estão sujeitos às mesmas restrições termodinâmicas. Existem vários ensembles, sendo que quatro são habitualmente os mais utilizados, e para cada um deles certas variáveis termodinâmicas são fixadas. No ensemble canónico (NVT) é mantido constante o número de partículas (N), o volume (V) e a temperatura (T); o ensemble microcanónico (NVE), em que N, V e a energia total (E) são mantidas constantes; o isotérmico-isobárico (NPT), em que N, T e a pressão (P) são

Capítulo I: Introdução

constantes; e o grande canônico (μVT), em que se mantém constante V , T e o potencial químico (μ). Mantendo estas funções de estado constantes, as restantes propriedades podem ser calculadas através das relações fundamentais da termodinâmica. Assim sendo, a escolha do ensemble a utilizar é feita com base naquilo que dará maior conveniência ao utilizador.

As equações de movimento de Newton (equação 9) podem ser usadas para estudar a trajetória de cada partícula no espaço. Para tal é preciso calcular a força resultante que atua em cada átomo e de seguida integrar as equações do movimento de forma a conhecer as suas posições em função do tempo.

$$F_i = m_i a_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad (9)$$

em que, m_i é a massa do átomo, a_i é a aceleração e $\frac{\partial^2 r_i}{\partial t^2}$ a segunda derivada da partícula na posição r . Uma vez que não existe uma solução analítica para estas equações, torna-se necessário resolvê-las numericamente. Para isso, é considerado um passo de integração finito (Δt), sendo as coordenadas do átomo obtidas no intervalo $t_0 + \Delta t$ (equação 10) e a velocidade no instante $t_0 + \Delta t$ (equação 11):

$$\vec{r}_i(t_0 + \Delta t) = \vec{r}_i(t_0) + \vec{v}_i(t_0) \cdot \Delta t + \frac{1}{2} \vec{a}_i(t_0) \cdot \Delta t^2 \quad (10)$$

$$\vec{v}_i(t_0 + \Delta t) = \vec{v}_i(t_0) + \vec{a}_i(t_0) \cdot \Delta t \quad (11)$$

De acordo com as novas posições atômicas, as forças são atualizadas e o processo repete-se até à obtenção do número de microestados desejados. Diferentes algoritmos podem ser utilizados para realizar a integração das equações, sendo o mais usado para as biomoléculas o algoritmo de Verlet [50]. É importante referir a aproximação que é feita ao se considerar que a força que atua no átomo i no intervalo $t_0 + \Delta t$ é constante e igual à força atuante no instante t_0 . Assim, estas trajetórias serão mais exatas quanto menor for o valor de Δt . Desta forma, a escolha do passo de integração (Δt) é fundamental na realização de uma simulação de DM.

3.3.2. Escolha do tempo de integração

Uma seleção cuidada do tempo de integração para ser usado na simulação de DM é crucial para o sucesso da simulação. Um passo de integração curto implica um aumento do tempo computacional necessário, o que diminui o espaço conformacional que pode ser estudado nesse tempo computacional. Um passo de integração largo pode levar à falha da simulação devido a uma incorreta integração das equações de movimento. Torna-se imprescindível

encontrar o balanço ideal entre economia e exatidão. Para as biomoléculas, este equilíbrio ditou um passo de integração de 1 fs, que corresponde a um valor dez vezes inferior à escala temporal associada ao movimento mais rápido, que, para este caso, é de 10 fs e está associado ao *stretching* de ligações que contêm átomos de hidrogénio (C-H). É, contudo, possível utilizar tempos de integração superiores sem comprometer a estabilidade da simulação. Para tal, é efetuada uma restrição das ligações que envolvam átomos de hidrogénio, o que faz com que o movimento seguinte mais rápido seja o da vibração das ligações envolvendo átomos pesados. Como este movimento é mais lento, torna-se possível utilizar um tempo de integração de 2 fs. Um método muito comum para fazer esta restrição é a utilização do algoritmo SHAKE[51].

3.3.3. Condições Periódicas de Fronteira

A água é o constituinte natural dos sistemas biológicos e como tal, quando pretendemos realizar simulações DM de sistemas biológicos, devemos procurar reproduzir da melhor forma o seu ambiente natural. Contudo, de forma a ser computacionalmente eficiente, uma simulação DM não pode reproduzir na íntegra o tamanho do sistema biológico. Os sistemas simulados podem atingir os 500 mil átomos contidos numa célula de simulação cúbica, paralelepípeda, octaédrica, entre outras. Um problema comum à célula de simulação é o chamado efeito de superfície. O efeito de superfície está relacionado com o facto de, numa caixa cúbica, as moléculas que se encontram nas faces da caixa estarem sujeitas a interações diferentes das moléculas do centro da caixa. Para ultrapassar esta situação, foi desenvolvida uma imposição designada por Condições Periódicas de Fronteira (CPF). Nas CPF a célula de simulação vai apenas representar a célula central de uma rede cúbica de réplicas suas (Figura I.10). A integração das equações de movimento é feita apenas para os átomos da célula central e os movimentos das moléculas nas células réplicas são mimetizados a partir dos da caixa de simulação. É importante considerar alguns fatores ao aplicar CPF: o aspeto fundamental ao aplicar CPF é que nenhuma molécula da célula de simulação esteja sujeita a vácuo dentro do intervalo de interação, evitando assim os artefactos de superfície; o sistema deve ser neutro, de forma a evitar cargas infinitas quando CPF são aplicadas, utilizando-se para tal contra-íões; a célula central deve ser suficientemente grande para que cada molécula não interatue de forma significativa com as suas réplicas.

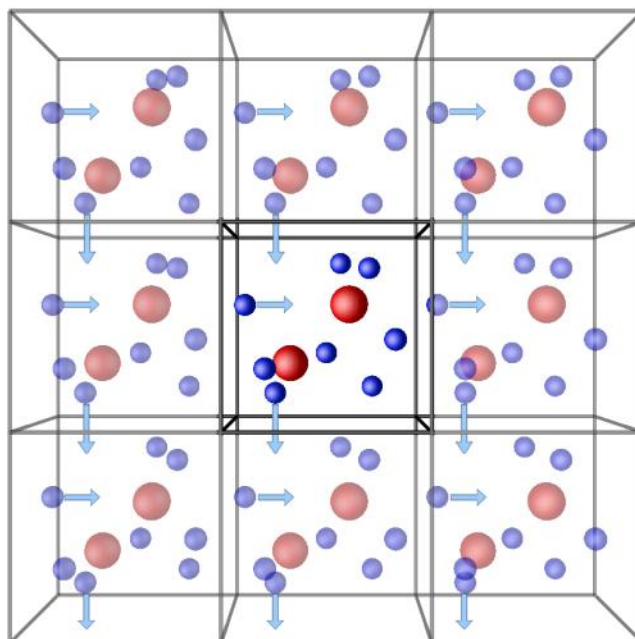


Figura I.10 - Representação esquemática de condições periódicas de fronteira. Em realce, no centro, encontra-se a célula de simulação e à sua volta as células réplica com o mesmo tipo de movimento da célula central.

3.3.4. Interações de curto e longo alcance

As interações intermoleculares podem ser divididas em interações de curto alcance (interações de Lennard-Jones) e interações de longo alcance (interações coulômbicas). As interações de longo alcance devem ser tratadas com especial atenção, pois a utilização de CPF aumenta infinitamente o alcance deste tipo de interações, uma vez que partículas de diferentes células podem interagir entre si. O cálculo dos dois termos não-ligados do CF, as interações eletrostáticas e as interações de van der Waals, é o mais exigente computacionalmente para cada passo de integração. Por isso ambas as interações são normalmente truncadas, limitando o número de interações entre pares. Até um determinado raio (r_{trunc}), as interações não-ligantes são consideradas. Com a utilização de CPF, o raio a considerar não deve ser maior do que metade do vetor da célula mais pequena. Depois são adicionados termos de correção para incluir a parte não considerada. O potencial de Lennard-Jones decai proporcionalmente a r^{-6} , podendo-se desprezar forças adicionais para r grandes. As interações de longo alcance variam com a distância proporcionalmente a r^{-1} . Uma das formas para lidar com as interações de longo alcance passa pela utilização de funções suavizantes, que vão anular gradualmente as interações ao longo de um intervalo de distâncias. Outra das formas, e a mais utilizada, é a utilização do método PME (Particle Mesh-Ewald). Neste método, calcula-se a soma das interações de longo alcance entre diferentes partículas, sendo para isso as interações divididas em curtas e longas pela adição de cargas gaussianas. Primeiro, cada carga pontual é

neutralizada para interações longas através da adição de uma nuvem de carga, com geometria esférica e densidade dada por uma função gaussiana centrada no átomo. Assim, apenas interações de curto alcance são determinadas. Depois o método adiciona um segundo conjunto de nuvens gaussianas, simétricas às primeiras, para cancelar o efeito da sua inclusão. O potencial é depois obtido a partir da equação de Poisson e pode ser calculado através de uma série de Fourier.

3.3.5. Controlo da Temperatura e Pressão

O carácter conservativo das equações de movimento, associado à tendência dos processos químicos e biológicos ocorrerem a temperaturas ou pressões constantes (ensembles NVT ou NPT), sendo por isso desejável realizar simulações DM nestas condições, tornou necessária a inclusão de métodos que regulassem a temperatura e a pressão do sistema na simulação. Devido ao carácter conservativo, a energia cinética iria variar simetricamente à variação da energia potencial, sendo esta dependente da posição de todas as partículas do sistema. Diferentes tipos de termóstatos foram desenvolvidos com o propósito de manter a temperatura do sistema constante, sendo os mais utilizados o termóstato de Nosé-Hoover[52], Berendsen[53] e Langevin[54, 55]. Para manter a pressão constante ao longo de uma simulação, é necessária que ocorram flutuações no volume. Alguns métodos já foram desenvolvidos para se executarem simulações a pressão constante, sendo os mais utilizados o baróstato de Berendsen e o pistão de Langevin[56].

3.4. Método de cálculo da energia livre

Interações entre proteínas e seus substratos têm um papel central em muitos processos biológicos, e como tal, é crucial que existam métodos capazes de calcular as variações de energia. O cálculo da energia livre de ligação entre dois estados, ΔG , tem sido muito utilizada na caracterização de sistemas biológicos, em especial, no cálculo da afinidade relativa de ligação de um inibidor ao seu substrato. Vários métodos foram desenvolvidos para este fim, com diferenças no poder computacional necessário e na qualidade dos resultados obtidos. Os métodos FEP (Free Energy Perturbation) e TI (Thermodynamic Integration) são os métodos, disponíveis atualmente, mais rigorosos para o cálculo da energia livre. Apesar de serem os métodos mais rigorosos são também os métodos mais exigentes computacionalmente, o que leva, por vezes, à sua substituição por métodos mais rápidos. Um destes métodos, MM-PBSA

Capítulo I: Introdução

(Molecular Mechanics-Poisson Boltzmann Surface Area) é utilizado como alternativa aos anteriores. Apesar de ser mais eficiente sob o ponto de vista computacional, é menos rigoroso.

3.4.1. MM-PBSA

O método MM-PBSA combina conceitos de mecânica molecular e modelos de solvente contínuo, sendo por isso um método híbrido para calcular a energia livre de ligação de diferentes conformações. Este método estima a energia livre do estado final diretamente, sem passar pelos estados intermédios, e é baseado no cálculo da energia livre a partir de contribuições entálpicas e entrópicas. Descreve a energia livre como uma soma da energia interna (*stretching, bending, torsion*), das interações eletrostáticas e de van der Waals, da energia livre de solvatação (G_{SOLV}) e da contribuição entrópica (S) (equação 12-14).

$$G_{molécula} = E_{MM} + G_{SOLV} - TS \quad (12)$$

$$E_{MM} = E_{interna} + E_{elec} + E_{vdw} \quad (13)$$

$$G_{SOLV} = G_{solv\ polar} + G_{solv\ apolar} \quad (14)$$

A energia livre do complexo pode ser calculada utilizando o ciclo termodinâmico que de seguida se apresenta:

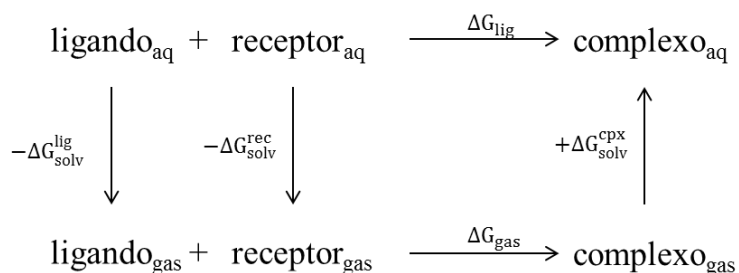


Figura I.11 - Ciclo termodinâmico utilizado no cálculo da energia livre do complexo.

A diferença na energia livre de ligação entre um complexo mutante e o respetivo complexo é dada pela seguinte expressão:

$$\Delta\Delta G_{ligação} = \Delta G_{ligação-mutante} - \Delta G_{ligação-wild\ type} \quad (15)$$

A energia livre de duas moléculas é a diferença entre a energia livre do complexo e a dos monómeros (recetor + ligando):

$$\Delta G_{\text{ligação-molécula}} = G_{\text{complexo}} - (G_{\text{recetor}} + G_{\text{ligando}}) \quad (16)$$

3.4.2. Solvatação

Uma vez que a grande maioria dos processos químicos ocorrem em solução, torna-se necessário contabilizar o efeito do solvente. Uma forma simples de contar com o efeito do solvente é o tratamento explícito de moléculas de água nas simulações de DM, em que as propriedades do solvente e do soluto são obtidas pela média de um grande número de configurações do sistema. Contudo, apesar do rápido crescimento no poder computacional e dos avanços nos métodos de simulação de DM, a procura exaustiva de tal espaço conformacional é muitas vezes impraticável para a maioria dos sistemas de interesse. Uma alternativa é a utilização de um modelo de solvatação contínuo. Neste modelo de solvatação o soluto é representado por um conjunto de átomos dentro de um baixo volume dielétrico, rodeado por um meio dielétrico elevado. Utilizando este método, as propriedades médias de solvatação da água podem ser obtidas sem a necessidade de medir as interações de milhares de moléculas de água explícita.

A energia livre de solvatação é a energia livre de transferência entre dois meios, e pode ser dividida em componentes apolar e polar, que são tratados energeticamente de forma independente:

$$\Delta G_{\text{solvatação}} = \Delta G_{\text{apolar}} + \Delta G_{\text{polar}} \quad (17)$$

O termo apolar inclui o custo energético da formação da cavidade, rearranjo do solvente, e interações soluto-solvente introduzidas quando o soluto não-carregado passa do vácuo para o solvente. Assim, este termo corresponde à energia livre de solvatação de uma molécula removida de todas as cargas (cargas parciais de cada átomo são tomadas como zero), seguindo da equação 17:

$$\Delta G_{\text{apolar}} = \Delta G_{\text{cavidade}} + \Delta G_{\text{dispersão}} \quad (18)$$

$\Delta G_{\text{dispersão}}$ é a energia das interações de van der Waals entre o solvente e o soluto, e $\Delta G_{\text{cavidade}}$ é um termo que inclui a penalização entrópica devida à reorganização das moléculas de solvente à volta do soluto e o trabalho realizado no solvente para criar a cavidade necessária para imergir o soluto. Ambos os termos são considerados proporcionais ao SASA (Solvent

Capítulo I: Introdução

Accessible Surface Area) do soluto, sendo o termo apolar estimado usando uma relação empírica:

$$\Delta G_{apolar} = \sigma A + \beta \quad (19)$$

em que A é o SASA estimado usando um software específico (como o programa Molsurf, integrado no pacote do AMBER [57]), que é baseado na ideia inicialmente desenvolvida por Michael Connolly [58]. σ e β são constantes empíricas com o valor 0,00542 kcal Å⁻² mol⁻¹ e 0,92 kcal mol⁻¹, respetivamente.

A parte polar da energia livre de solvatação descreve a energia livre relativa à carga do soluto no meio dielétrico, podendo ser estimada por diversas vias analíticas que aproximam a solução, computacionalmente exigente, da equação de Poisson-Boltzmann. Dois tipos de aproximação são vulgarmente utilizadas, a aproximação Poisson-Boltzmann (PB) e a aproximação Generalized Born (GB), que tratam o soluto como um conjunto de cargas pontuais num meio dielétrico baixo e o solvente circundante como um meio dielétrico elevado. O modelo PB é genericamente considerado como sendo mais preciso, sendo utilizado como referência para modelos GB. O modelo GB tem encontrado maior aplicação em simulações de DM, uma vez que é mais eficiente do ponto de vista computacional. Um dos programas mais utilizados para resolver a equação de PB para um sistema proteína-solvente é o DelPhi [59]. A equação mais usual de PB é dada pela seguinte expressão:

$$\nabla(\epsilon(\vec{r})\nabla\phi(\vec{r})) = -4\pi\rho(\vec{r}) \quad (20)$$

em que $\phi(\vec{r})$ descreve o potencial eletrostático num meio com densidade de carga $\rho(\vec{r})$ e um campo dielétrico $\epsilon(\vec{r})$. Esta é uma equação diferencial parcial elíptica de segunda ordem, que pode ser resolvida analiticamente para solutos de forma regular e, numericamente, para outros de forma irregular. Geralmente, a energia de solvatação eletrostática é obtida pela diferença entre estados adequados, em energias de reação.

O DelPhi apresenta uma forma alternativa de calcular a energia de solvatação eletrostática, baseada no conceito de cargas induzidas, a partir de um cálculo de Coulomb entre as cargas induzidas e as cargas reais como se estivessem em vácuo. No DelPhi, a energia de solvatação eletrostática é calculada pelo Método da Diferença Finita, que é baseado na discretização do espaço numa grelha uniforme. A definição da fronteira entre o interior e exterior de uma dada molécula é de grande importância, uma vez que esta é a região em que a constante dielétrica altera entre o seu valor interno e o valor superior associado à maioria dos solventes. Esta

fronteira corresponde à superfície molecular definida pelo contacto entre a superfície de van der Waals de uma molécula e a superfície de uma sonda esférica que representa o solvente. Um parâmetro fundamental nestes cálculos é pois, a constante dielétrica interna.

A solução numérica da equação de PB tem encontrado múltiplas aplicações em campos como a Química e a Biologia. Muitas das propriedades associadas a moléculas em solução são eletrostáticas e, em muitos casos, a aproximação a um meio contínuo implícito na equação de PB tem-se revelado precisa no tratamento destes fenómenos. Diferentes tipos de técnicas e programas têm sido utilizados para resolver esta equação. A equação de PB é maioritariamente aplicada na sua forma linear, como descrita na teoria de Deybe-Huckel.

A equação não-linear de PB para uma macromolécula numa solução de sal 1:1 é dada por:

$$\nabla \cdot [\varepsilon(r) \nabla \cdot \phi(r)] - \varepsilon k^2 \sinh[\phi(r)] + \frac{4\pi e \rho^f(r)}{kT} = 0 \quad (22)$$

em que $\phi(r)$ é o potencial eletrostático adimensional em unidades kT/e em que k é a constante de Boltzmann, T é a temperatura absoluta e e é a carga do protão. Adicionalmente, ε é a constante dielétrica e ρ^f é a densidade de carga. O termo $k^2 = 1/\lambda^2 = 8\pi e^2 I / \varepsilon kT$, em que λ é o comprimento de Debye e I a força iónica da solução. As quantidades ϕ , ε , k e ρ são todas funções do vector de posição r no frame de referência centrado numa macromolécula fixa [60].

Diferentes soluções numéricas para a equação não-linear de PB estão disponíveis e já demonstraram a sua precisão na contabilização de efeitos do sal em interações eletrostáticas envolvendo DNA [60]. Atualmente existe um número apreciável de formalismos desenvolvidos no contexto da equação não-linear de PB que permitem atribuir diferentes constantes dielétricas a diferentes regiões do espaço conformacional, bem como diferentes concentrações iónicas. Estas expressões foram incluídas sob a forma de algoritmos no programa DelPhi, que assim nos providencia soluções para a equação não-linear de PB para o caso de moléculas polares e carregadas, com constantes dielétricas internas variáveis, envolvidas num meio que também ele pode apresentar diferentes constantes dielétricas e concentrações variáveis de iões [61, 62].

3.4.3. Constante dieletrica interna

Um sistema formado por proteínas é altamente polar, uma vez que é formado por aminoácidos com carga e dipolos. Contudo, a distribuição de resíduos polares e carregados não é homogénea. Estes resíduos tendem a aparecer à superfície das moléculas enquanto que

Capítulo I: Introdução

os resíduos hidrofóbicos encontram-se geralmente no centro da proteína. Tal não se verifica para todos os sistemas, como por exemplo as proteínas membranares, que apresentam frequentemente resíduos polares e carregados no seu centro. Isto indica que a polaridade de macromoléculas biológicas varia desde o centro até à superfície, dependendo da estrutura e composição em aminoácidos da macromolécula.

A constante dielétrica interna ($\epsilon_{\text{interna}}$) funciona como um meio de contabilizar respostas à ação de um campo elétrico não tratado explicitamente. Esta resposta depende da constituição dos resíduos e por isso diferentes regiões da proteína devem ter diferentes constantes dielétricas. Cada aminoácido é considerado como sendo um objeto formado pelos respetivos átomos e por cargas parciais, descrito como um meio homogéneo com uma constante dielétrica particular. As constantes dielétricas mais baixas são atribuídas a aminoácidos com baixa carga parcial (resíduos hidrofóbicos) e pouco flexíveis, enquanto que aminoácidos polares e carregados localizados na superfície e contendo elevados dipolos elétricos vão apresentar constantes dielétricas mais elevadas. Geralmente, em cálculos PB, a proteína é tratada como um meio homogéneo com constante dielétrica entre 2 (reflectindo a polarizabilidade eletrónica) e 4. Contudo, outros trabalhos, apresentam variações e constantes dielétricas óptimas mais elevadas, desde 1-36 até mais de 80, dependendo da concentração do sal [63].

3.5. Método computacional Alanine-Scanning Mutagenesis

O método Computacional *Alanine-Scanning Mutagenesis* (ASM) é um dos métodos mais utilizados na análise de interfaces e na deteção de HS. Este método permite aferir do papel desempenhado pelos grupos funcionais das cadeias laterais dos aminoácidos em posições específicas, bem como da contribuição individual de cada cadeia lateral para a ligação, uma vez que a substituição por alanina remove os átomos da cadeia lateral localizados depois do carbono- β , sem com isso introduzir liberdade conformacional. O método envolve a mutação do resíduo pretendido para uma alanina e cálculo da energia livre de ligação. A mutação para glicina também anula o efeito da cadeia lateral, mas pode introduzir flexibilidade conformacional no *backbone*, e por isso não é tão utilizado. Neste trabalho, considerou-se o valor de 2,0 kcal/mol, referente à energia livre de ligação, para definir um HS. Resíduos com energia livre de ligação menor do que 2,0 kcal/mol foram considerados null-spots (NS).

Assim, foi proposto um método em que se utilizavam diferentes constantes dielétricas internas para diferentes mutações, com valores dependentes do tipo de aminoácido que era mutado para alanina. Desta forma, para os diferentes grupos de aminoácidos encontrados em proteínas, nomeadamente, apolares (valina, alanina, leucina, isoleucina, fenilalanina, prolina, glicina, metionina e triptofano), polares e neutros (asparagina, glutamina, cisteína, tirosina, serina e treonina), ácidos e carregados (ácido aspártico e ácido glutâmico), básicos e carregados (lisina, arginina e histidina) diferentes valores de constantes dielétricas seriam usados de forma a obter a melhor correlação com os resultados experimentais. Apesar de se ter este conceito como base, neste trabalho iremos testar outras constantes de forma a averiguar aquelas que melhor se enquadram neste tipo de interfaces.

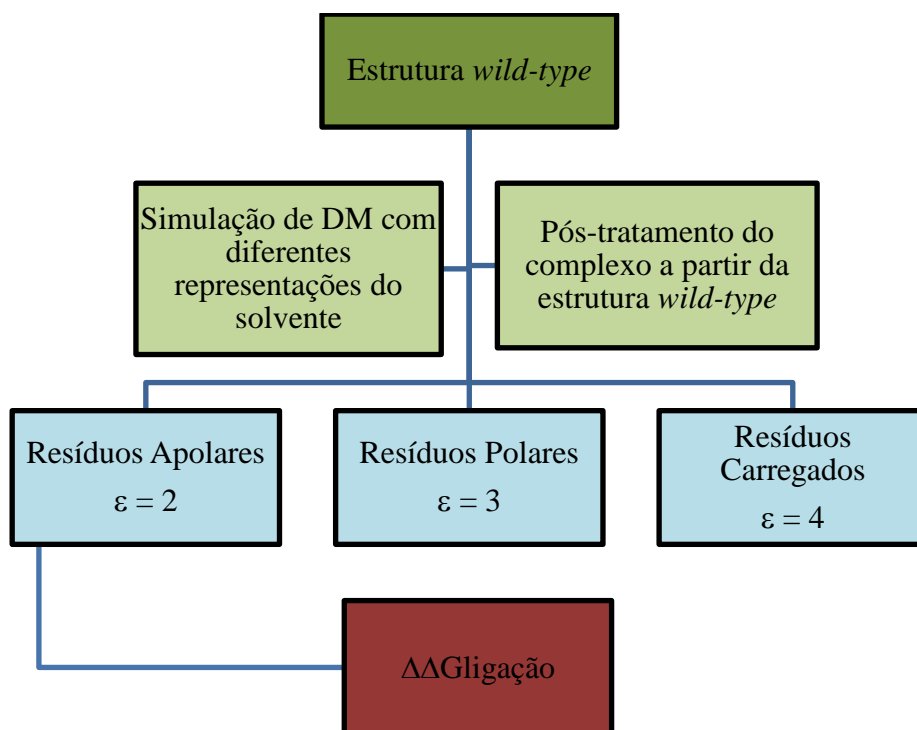


Figura I.12 – Resumo da metodologia computacional ASM aplicada ao estudo de sistemas proteína-proteína

Capítulo II. Metodología

CAPÍTULO II: METODOLOGIA

Prefácio

Para uma melhor estruturação deste trabalho e de forma a proporcionar uma leitura mais simples do mesmo, resolveu-se dividir o capítulo da metodologia em duas partes: Parte 1 – referente ao estudo da mutagénese em interfaces proteína-DNA; Parte 2 – na qual se descreve o protocolo utilizado no estudo do mecanismo de inibição do Tenofovir, incluindo a preparação dos sistemas para as simulações de DM, a própria simulação de DM e a análise dos resultados.

1. MUTAGÉNESE EM INTERFACES PROTEÍNA-DNA

1.1. Preparação dos sistemas

Uma vez que não existe grande informação de ASM para interfaces proteína-DNA, começou-se por fazer uma pesquisa em bases de dados por sistemas proteína-DNA com valores experimentais de variação da energia livre, e para os quais houvesse mutação de aminoácidos para alanina. Uma das bases de dados utilizada foi a ProNIT [64]. A estrutura cristalográfica dos complexos escolhidos foi obtida no RCSB Protein Data Bank (PDB). O estado de protonação dos diferentes resíduos nos vários complexos foi determinado usando a metodologia PROPKA[65-67], que se baseia em ligações de hidrogénio favoráveis, energias locais e valores de pK_a . O programa *leap* (incluído no pacote de programas do AMBER) foi utilizado para criar os ficheiros de *input* da simulação de DM. O campo de forças utilizado foi o *ff99SB* (AMBER).

Utilizou-se o *leap* para solvatar cada sistema com uma caixa de moléculas de água do tipo TIP3P [68], que se prolongavam por uma distância de 10 Å desde a borda da caixa até aos átomos da proteína. Quando se seleciona um modelo para adição de solvente a um sistema, é importante que este seja compatível com o CF utilizado. Apesar de existirem mais modelos, o TIP3P é provavelmente o mais utilizado devido às boas propriedades dinâmicas e termodinâmicas que proporciona para a maioria dos CF. Uma quantidade apropriada de contra-íões sódio foi utilizada para neutralizar o sistema. Para finalizar, foram criados os ficheiros de *input* para a simulação de DM: ficheiro de topologia (*.prmtop*) e ficheiro de

Capítulo II: Metodologia

coordenadas (*.inpcrd*). Um resumo das características dos sistemas submetidos a DM pode ser consultado na Tabela II.1.

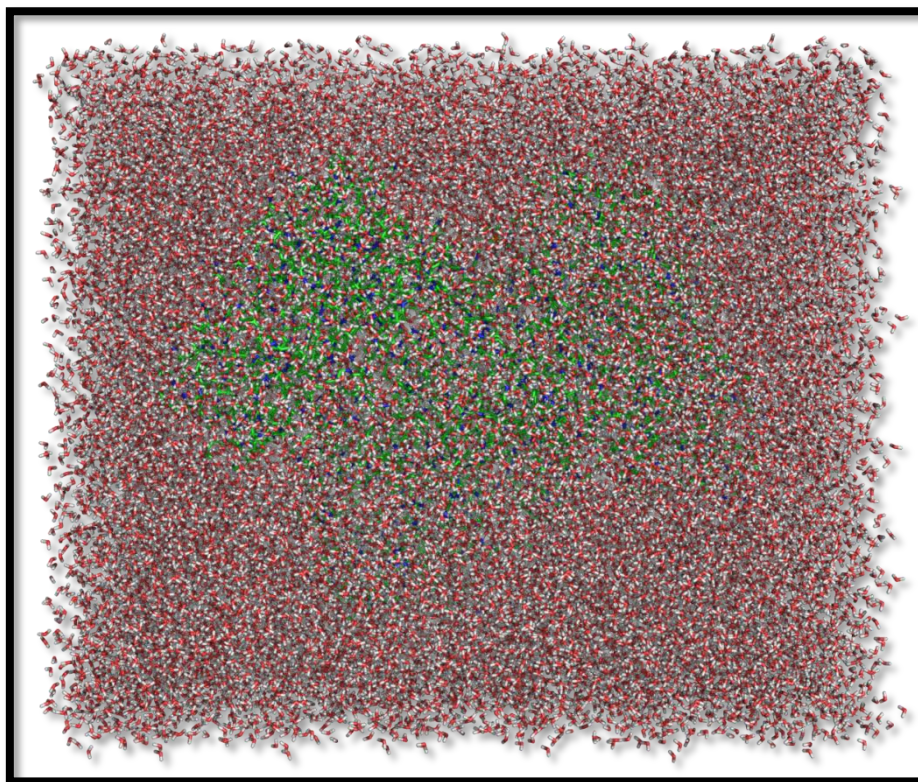


Figura II.1 - Representação de um sistema solvatado com uma caixa de águas do tipo TIP3P.

Tabela II.1 - Composição em termos de resíduos (número de aminoácidos, ácidos nucleicos, águas e iões) e átomos dos sistemas proteína-DNA em estudo.

Complexo	Resíduos					Átomos	Ref.
	AA	Ácidos Nucleicos	Águas	Iões	Total		
1MNM	319	52	26.214	21 Na ⁺	26.607	85.720	[69]
1BDT	204	44	13.310	28 Na ⁺	13.587	44.809	[70]
1MSE	104	22	7.213	8 Na ⁺	7.348	24.161	[71]
1QRV	144	19	11.224	12 Na ⁺	11.400	36.660	[72]
1B3T	294	35	13.325	26 Na ⁺	13.681	45.663	[73]

Apresenta-se de seguida uma representação simples dos cinco complexos em estudo: 1MNM (Fator Transcrição – *Alpha / Beta*); 1BDT (Fator Transcrição – *Ribbon / Helix / Helix*); 1MSE (Proteína de Suporte); 1B3T (Proteína de Suporte); 1QRV (Fator Transcrição – *Alpha Helix*) (Figura II.2).

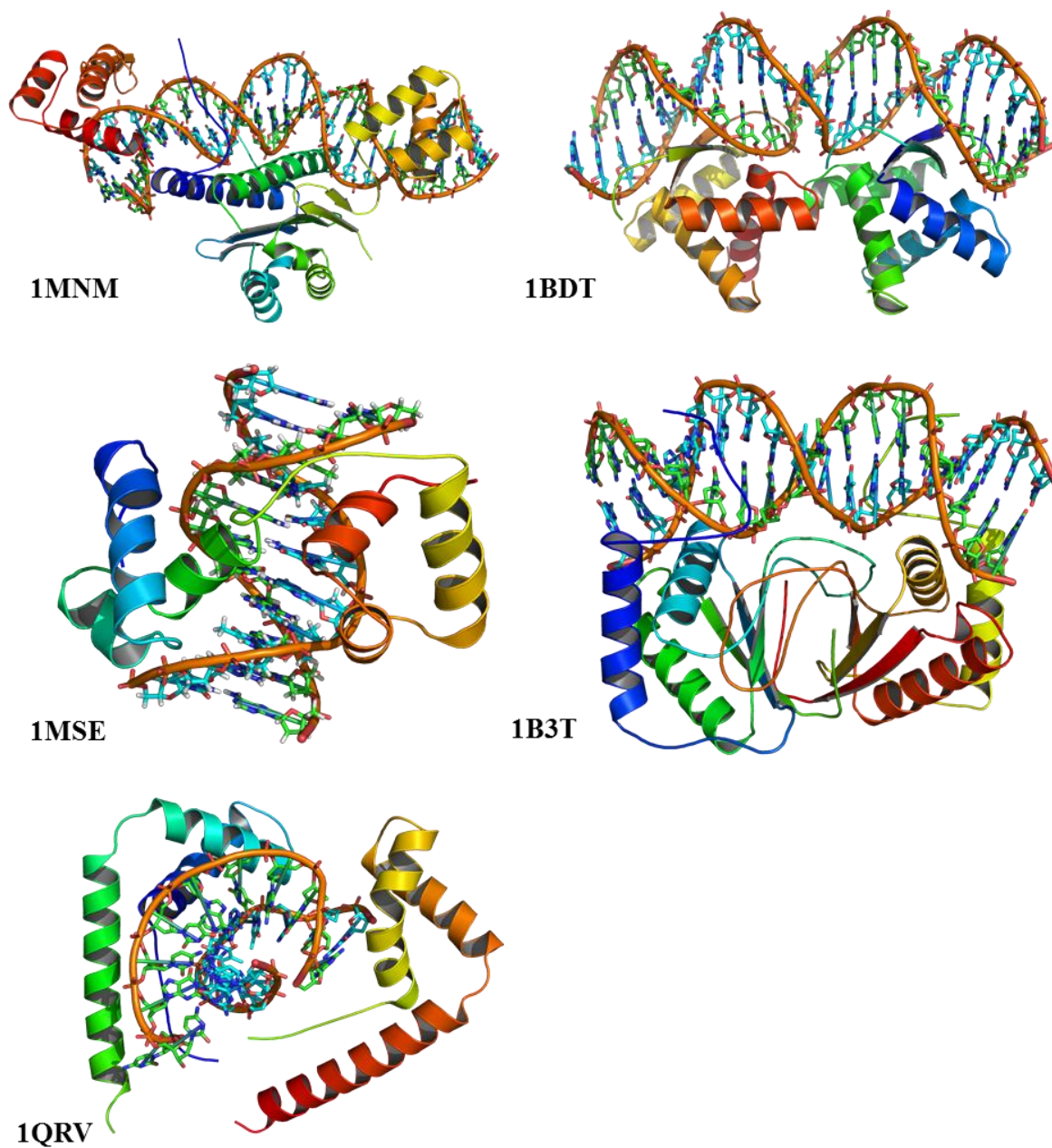


Figura II.2 - Representação dos cinco complexos proteína-DNA estudados neste trabalho.

Capítulo II: Metodologia

1.2. Simulação de Dinâmica Molecular

Após a obtenção dos parâmetros de *input* para os vários sistemas começou-se a preparar as simulações de DM, precedidas por uma minimização que pode ser dividida em duas etapas: uma etapa de aquecimento e uma etapa de produção. A execução de uma etapa de minimização é aconselhável quando se realizam simulações de DM de forma a eliminar possíveis maus contactos existentes na estrutura cristalográfica ou na interação com o solvente. A minimização foi realizada com o módulo SANDER incluído no pacote AMBER9.

A etapa de produção foi realizada, também ela, com o módulo SANDER. Os sistemas foram sujeitos a 2 ns de aquecimento, em que gradualmente se aumentou a temperatura de 0 a 300K (ensemble NVT) seguido de 8 ns de produção em ensemble NPT. O algoritmo de Langevin foi utilizado para regular a temperatura do sistema. Condições periódicas de fronteira foram aplicadas usando o método PME para tratar interações eletrostáticas de longo alcance, sendo as interações não-ligantes truncadas com um raio de 10Å. Em todas as simulações, as ligações envolvendo átomos de hidrogénio foram restringidas pelo logaritmo SHAKE, permitindo a utilização de um tempo de integração de 2-fs. Realizámos simulações de 10 ns para todos os sistemas, sendo as coordenadas gravadas a cada 2 ps.

As simulações de DM em solvente implícito foram realizadas com o modelo de solvatação de Born, implementado no SANDER, mais propriamente o IGB5. Tal como em solvente explícito, realizaram-se simulações de 10 ns.

1.3. Alanine-Scanning Mutagenesis

O programa MM-PBSA integrado no AMBER9 foi utilizado para o cálculo da diferença na energia livre de ligação ($\Delta\Delta G$) após mutação do resíduo para alanina. Como referido anteriormente, a expressão para o cálculo da energia livre contém termos da energia interna, interações eletrostáticas e de van der Waals, energia livre de solvatação polar, energia livre de solvatação não-polar e contribuições entrópicas.

$$G_{\text{molécula}} = E_{\text{interna}} + E_{\text{eletrostática}} + E_{\text{vdW}} + G_{\text{solv polar}} + G_{\text{solv apolar}} - TS \quad (23)$$

No cálculo de energias relativas entre complexos próximos assume-se que o termo entrópico da equação 23 é negligenciável, pois as contribuições parciais de cada complexo acabam por se anular. Os três primeiros termos da equação 23 foram calculados sem qualquer corte. O termo da energia livre de solvatação polar foi calculado através da resolução da equação linear de Poisson-Boltzmann com o programa DelPhi. Neste método, o sistema é modelado

como um meio contínuo dielétrico de baixa polaridade envolvido num meio dielétrico de alta polaridade. Utilizou-se um conjunto de valores como parâmetros do DelPhi que, em estudos anteriores, demonstraram ter uma boa relação precisão / tempo de computação. Assim sendo, utilizou-se um valor de 2,5 grelhas/Å de escala; 0,001 kT/c como critério de convergência; 90% de ocupação da grelha. A fronteira dielétrica foi considerada como a superfície molecular definida por uma esfera de raio 1,4 Å e por esferas centradas em cada átomo. Tendo em conta a importância dos valores das constantes dielétricas usadas, demonstrado no estudo por ASM de sistemas proteína-proteína, neste trabalho utilizaram-se valores de 1 a 10 para a constante dielétrica de forma a verificar quais seriam as mais corretas para simular o rearranjo após mutação para alanina. Uma vez que a mutação sistemática de resíduos numa interface é trabalhosa e feita de forma lenta, recorreu-se a um plugin para o VMD designado CompASM [74]. Este plugin oferece-nos uma interface gráfica de fácil utilização para a preparação dos ficheiros de *input*, cálculo e análise dos resultados.

Neste trabalho aplicou-se o método ASM a um conjunto de sessenta mutações, oriundas dos cinco complexos proteína-DNA anteriormente referidos. Na tabela seguinte apresentam-se as várias mutações analisadas, bem como os seus valores experimentais de energia livre de ligação.

Tabela II.2 - Resumo das mutações em estudo e seus valores de energia livre experimentais.

Proteína	#AA PDB	AA Mutado	$\Delta\Delta G_{exp}$ / kcal mol ⁻¹	Referência
1BDT	4	M	1,10	[75]
1BDT	5	S	1,30	
1BDT	6	K	1,30	
1BDT	7	M	1,90	
1BDT	9	Q	1,80	
1BDT	11	N	2,00	
1BDT	13	R	6,00	
1BDT	23	R	0,20	
1BDT	29	N	-1,00	
1BDT	31	R	-0,30	
1BDT	32	S	-2,30	
1BDT	33	V	-2,30	
1BDT	34	N	3,20	
1BDT	35	S	-0,20	
1BDT	39	Q	-0,40	
1B3T	469	R	3,41	[76]
1B3T	518	Y	2,62	
1B3T	522	R	4,40	
1MSE	116	S	0,06	[77, 78]
1MSE	139	N	0,60	

Capítulo II: Metodologia

1MSE	141	E	-0,10	
1MSE	187	S	0,10	
1QRV	9	L	0,02	
1QRV	13	M	1,20	[79]
1QRV	32	V	-0,30	
1MNM	16	K	0,54	
1MNM	17	E	0,24	
1MNM	18	R	0,28	
1MNM	19	R	0,81	
1MNM	20	K	0,64	
1MNM	21	I	0,88	
1MNM	22	E	0,20	
1MNM	23	I	0,86	
1MNM	24	K	-0,13	
1MNM	25	F	4,05	
1MNM	26	I	4,05	
1MNM	27	E	0,41	
1MNM	28	N	1,75	
1MNM	29	K	-0,13	
1MNM	32	R	2,76	
1MNM	33	H	-0,81	
1MNM	34	V	0,58	
1MNM	35	T	0,64	[80]
1MNM	36	F	4,05	
1MNM	37	S	0,43	
1MNM	38	K	4,05	
1MNM	39	R	4,05	
1MNM	40	K	2,83	
1MNM	41	H	-1,24	
1MNM	43	I	3,24	
1MNM	44	M	4,05	
1MNM	45	K	4,05	
1MNM	46	K	4,05	
1MNM	48	F	-0,13	
1MNM	49	E	2,27	
1MNM	51	S	0	
1MNM	52	V	0,11	
1MNM	53	L	4,05	
1MNM	54	T	4,05	
1MNM	66	T	0,64	

1.4. Análise dos resultados

1.4.1. RMSD

A análise dos sistemas obtidos por DM começou por ser feita com o cálculo do RMSD (Root Mean Square deviation). A análise por RMSD permite comparar duas conformações de um dado conjunto de pontos, fornecendo uma medida da diferença na posição desses mesmos conjuntos. Quando maior for o valor de RMSD maior será a diferença entre as estruturas. O nosso objetivo ao realizar este cálculo foi verificar se as estruturas geradas por DM já se encontram em equilíbrio ou não, tendo para isso utilizado a estrutura inicial do sistema como referência. Para fazer o cálculo de RMSD utilizou-se o programa *PTRAJ* incluído no pacote AMBER9. Neste trabalho, procedeu-se à medição de dois tipos de RMSD para cada um dos sistemas em análise: proteína e DNA.

1.4.2. RDF

A medição de RDFs é de extrema importância quando lidamos com sistemas contendo moléculas de solvente (solvente explícito), como são o caso das simulações de DM realizadas no contexto deste trabalho, uma vez que permite caracterizar a interação entre um soluto e as moléculas de um solvente. As águas localizadas à volta de complexos formados por proteínas podem ser divididas em três categorias: i) água que envolve a estrutura da proteína, livre de se mover e assistir na difusão da proteína relativamente a outras moléculas através de movimentos casuais em solução; ii) água de hidratação na superfície da proteína; e iii) moléculas individuais de águas ligadas, que formam pontes de hidrogénio com resíduos carregados ou polares, estabilizando a estrutura da proteína. De uma forma geral, a função de distribuição radial dá a densidade de partículas de solvente a uma distância r das partículas do soluto.

2. MECANISMO DE INIBIÇÃO DO TENOFOVIR

2.1. Preparação dos sistemas

No presente trabalho, fez-se o estudo de oito complexos, no total, formados pela HIV-1 RT e pelo TDF ou pelo ligando natural. A estrutura cristalográfica tridimensional destes complexos foi obtida no RCSB Protein Data Bank (Tabela II.3). O estado de protonação dos diferentes resíduos nos vários complexos foi determinado pela metodologia PROPKA[65-67]. Um dos estados de protonação fisiológicos mais vezes afetado era o de aspartatos (ASP), que eram alterados para ASH, devido ao enriquecimento em resíduos polares na sua vizinhança. A parametrização dos vários ligandos foi efetuada com recurso aos programas *Antechamber*[81] e *Gaussian09*[82].

O *Antechamber* constitui um pacote de programas incluído no AMBER que permite o estudo de moléculas através de MM. Este programa tem a capacidade de atribuir cargas atômicas, tipos de átomos e tipos de ligações. Suporta vários métodos de atribuição de cargas, desde o RESP, AM1-BCC, CM1, etc., e é compatível com o GAFF. No caso de, mesmo com a utilização do GAFF, faltarem parâmetros de CF para algumas moléculas, um programa designado *parmchk* pode ser aplicado para encontrar os parâmetros de CF em falta, sugerir substituições com base em parâmetros similares ou calcular novos parâmetros com base em regras empíricas. A sua principal função, e aquilo para que é mais utilizado, é gerar a topologia do sistema.

Gaussian09 é a versão atual do famoso programa da série Gaussian[®], inicialmente publicado na década de 70 por John Pople. Este programa tornou-se uma ferramenta essencial para químicos e bioquímicos computacionais uma vez que, partindo das leis fundamentais da mecânica quântica, permite prever energias, estruturas moleculares, frequências vibracionais, propriedades moleculares de moléculas e reações numa variedade de sistemas químicos.

Tabela II.3 - Descrição dos vários sistemas em estudo. Informação do tipo de ligando e o seu local de ligação no complexo, bem como a resolução da estrutura cristalográfica.

PDB ID	Ligando	Local de ligação	Resolução / Å	Referência
1T05	TDF	N-site	3,00	[83]
1RTD	dTTP	N-site	3,20	[19]
1T03A	TDF	P-site	3,10	[83]
1T03B	TDF	P-site	3,10	[83]
1R0A	dATP	P-site	2,80	[27]
3JSM	TDF	N-site	3,00	[32]
3JYT	dATP	N-site	3,30	[32]
1N5Y	AZT	P-site	3,10	[84]

Utilizou-se o método quântico Hartree-Fock (HF) em conjunto com o programa *Gaussian09* no tratamento do ligando dos vários sistemas estudados. O HF é um dos métodos mais antigos e mais utilizados por químicos computacionais. Neste trabalho, utilizou-se o método HF com a base 6-31G (d) e método RESP para a atribuição das cargas atómicas. O RESP foi escolhido devido à sua compatibilidade com a maioria dos CF do AMBER (*ff99SB*) e com o GAFF, que foram os CF utilizados neste trabalho. A criação dos ficheiros de *input* para a DM decorreu com base no método descrito anteriormente (secção 1.1.), em que se utilizou o programa *leap* para adicionar a caixa de águas, os contra-íões e por fim criar os ficheiros de topologia e coordenadas. Apresenta-se de seguida uma tabela com algumas das características dos sistemas que foram submetidos a simulações DM (Tabela II.4).

Capítulo II: Metodologia

Tabela II.4 - Descrição da constituição dos vários sistemas submetidos a DM. Apresenta-se a sua constituição em aminoácidos, ácidos nucleicos, moléculas de água e iões.

Complexo	Resíduos					Átomos Total
	Aminoácidos	Ácidos Nucleicos	Águas	Iões	Total	
1T05	967	44	37.424	21 Na ⁺	38.461	129.684
1RTD	967	45	62.300	26 Na ⁺	63.348	204.382
1T03A	986	41	41.972	20 Na ⁺	43.024	143.584
1T03B	987	42	41.971	22 Na ⁺	43.025	143.581
1R0A	987	42	41.633	28 Na ⁺	42.694	142.602
3JSM	967	43	37.360	24 Na ⁺	38.399	129.473
3JYT	967	43	36.803	33 Na ⁺	37.848	127.806
1N5Y	987	40	41.825	27 Na ⁺	42.880	143.082

2.2. Simulações de Dinâmica Molecular

As estruturas cristalográficas de raio-X disponíveis no RCSB Protein Data Bank, e as estruturas tridimensionais (como referido anteriormente) que fornecem a base de conhecimento inicial para este trabalho, são estáticas e não mostram as várias modificações conformacionais que ocorrem no sistema ao longo do tempo. A correta compreensão dos fenómenos que ocorrem na HIV-1 RT, bem como a compreensão da adaptabilidade estrutural causada pela incorporação de um inibidor, podem beneficiar bastante da utilização de métodos computacionais capazes de gerar informações baseadas, não numa estrutura estática, mas num ensemble gerado por simulações de DM.

Na realização das simulações de DM dos sistemas indicados anteriormente, utilizou-se o procedimento descrito na secção 1.2., com a diferença de se terem prolongado as simulações até aos 18 ns.

2.3. Análise dos resultados

Na análise dos sistemas para além dos cálculos referidos anteriormente, realizaram-se outros que, de certa forma, nos dariam informações sobre o processo que se estaria a estudar. Entre estes, inclui-se a medição de distâncias entre aminoácidos chave ou entre aminoácidos e o inibidor; o cálculo de ângulos diedros para o ligando como forma de caracterizar a sua mobilidade conformacional; o estudo da interface que rodeia o inibidor, medição de ligações de hidrogénio e pontes iónicas, entre outros. Este último ponto, em que se estudou o ambiente que rodeava o ligando ao longo da simulação, foi efetuado com o programa VMD[85] e com scripts desenvolvidos para o tratamento dos dados obtidos, e permitiu-nos obter informações sobre o tipo de resíduos que ao longo da simulação faziam parte da esfera de coordenação do ligando (se eram resíduos carregados positiva ou negativamente, se eram polares ou apolares, águas, e o seu respetivo tempo de permanência). Programa essencial nesta análise, para além do já referido PTRAJ, é o programa VMD. O VMD é uma ferramenta essencial no estudo computacional de qualquer sistema biológico devido à interface gráfica de visualização que proporciona, seja para pdb's ou dinâmicas, ou mesmo funcionando como programa de cálculo uma vez que cada vez mais são desenvolvidos *plugins* especificamente para serem incorporados neste programa[86].

Capítulo III. Resultados e Discussão

CAPÍTULO III: RESULTADOS E DISCUSSÃO

1. MUTAGÊNESE EM INTERFACES PROTEÍNA-DNA

1.1. Composição das interfaces

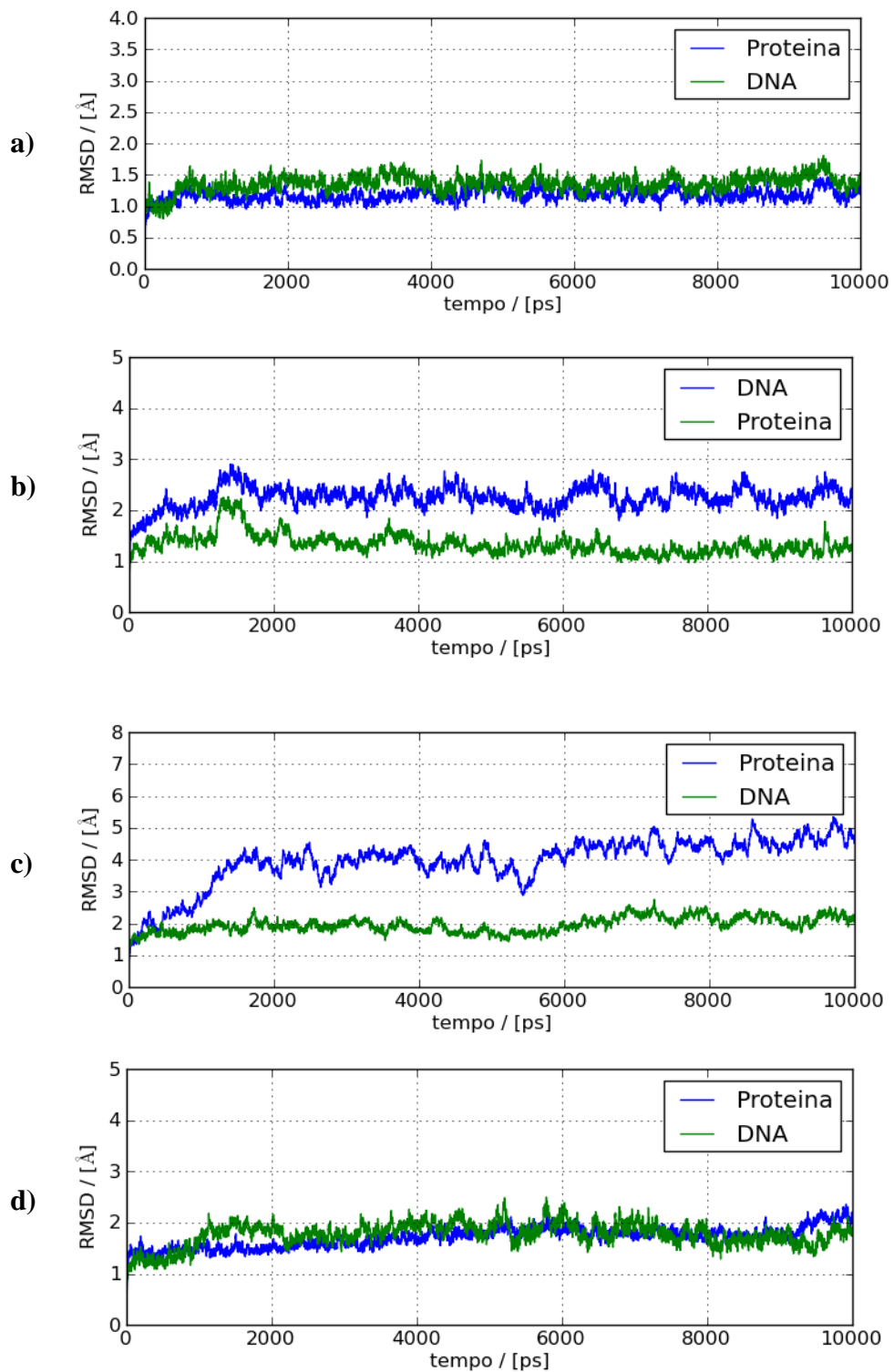
Das sessenta mutações analisadas neste trabalho, em termos dos diferentes grupos de aminoácidos, temos a seguinte distribuição: Glu (8 %), Phe (5 %), His (3 %), Ile (6 %), Lys (15 %), Leu (3 %), Met (8 %), Asn (8 %), Gln (5 %), Arg (15 %), Ser (11 %), Thr (5 %), Val (6 %) e Tyr (2 %). Deste grupo de resíduos, 40% são carregados, 31% são polares e 29% são apolares; 33% são HS e 67% são NS.

Estes resíduos, cuja escolha foi limitada à existência de valores de energia livre de ligação experimentais após mutação para alanina, representam o caráter característico de uma interface proteína-DNA com a elevada ocorrência de aminoácidos carregados, em especial a Lisina e a Arginina [15].

1.2. Estabilidade dos sistemas proteína-DNA: cálculo de RMSD

A avaliação da estabilidade dos sistemas resultantes das várias simulações de DM foi efetuada através do cálculo da variação de RMSD para a proteína e para o DNA. Para o cálculo considerou-se o backbone da proteína e os átomos P, O3*,C3*,C4*,C5* e O5* do DNA. Pela análise da variação de RMSD obtida para os cinco sistemas estudados (Figura III.1) é possível confirmar a estabilidade dos mesmos, uma vez que o RMSD tanto para a proteína como para o DNA é constante ao longo dos 10 ns de simulação, confirmando a manutenção da estrutura inicial ao longo da simulação. Existe, contudo, uma pequena exceção no caso do sistema 1MNM, em que se verifica uma maior variação da componente proteica.

Capítulo III: Resultados e Discussão



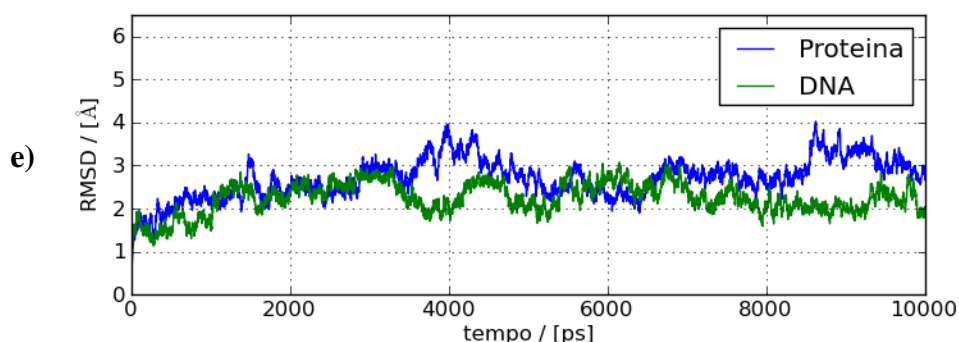


Figura III.1 - Representação da variação de RMSD da componente proteica e DNA para os sistemas 1B3T (a), 1BDT (b), 1MNM (c), 1MSE (d) e 1QRV (e).

1.3. Importância do tempo de simulação na determinação de $\Delta\Delta G$

Inicialmente resolveu-se averiguar a importância do tempo de simulação no cálculo da energia livre de ligação para cada uma das mutações em estudo neste trabalho. Para tal calculou-se a energia livre de ligação para um conjunto de 25 estruturas geradas dos últimos 2 ns das simulações de DM efectuadas em solvente explícito. De seguida, o mesmo tipo de cálculo foi efetuado para um conjunto de 75 estruturas geradas dos últimos 6 ns das simulações de DM. De forma a facilitar a discussão dos resultados e o próprio seguimento deste manuscrito resolveu-se apresentar, de seguida, apenas os resultados referentes às constantes dieléctricas 3 e 7 (Tabela III.1). A totalidade dos resultados pode ser consultada no Capítulo VI.

Tabela III.1 – Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns e 6 ns em solvente explícito. Constantes dieléctricas 3 e 7.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$			
			2 ns		6 ns	
			$\epsilon 3$	$\epsilon 7$	$\epsilon 3$	$\epsilon 7$
1BDT	M	1,10	4,68	4,92	4,31	4,69
1BDT	S	1,30	2,18	0,67	2,41	0,90
1BDT	K	1,30	76,46	34,94	77,09	35,22
1BDT	M	1,90	-0,36	-0,06	-0,26	-0,01
1BDT	Q	1,80	4,42	4,22	4,77	4,36
1BDT	N	2,00	3,40	3,28	3,90	3,53
1BDT	R	6,00	112,64	53,32	116,26	55,66
1BDT	R	0,20	95,61	43,84	96,08	44,05
1BDT	N	-1,00	1,42	0,66	1,28	0,60
1BDT	R	-0,30	71,15	32,17	71,04	32,12
1BDT	S	-2,30	4,14	1,77	4,57	2,12
1BDT	V	-2,30	3,71	1,93	3,70	1,99

Capítulo III: Resultados e Discussão

1BDT	N	3,20	-1,08	-0,16	-1,13	-0,16
1BDT	S	-0,20	0,90	0,58	1,06	0,62
1BDT	Q	-0,40	0,08	0,05	-0,16	-0,08
1B3T	R	3,41	110,52	57,20	111,64	57,60
1B3T	Y	2,62	8,08	7,15	7,46	6,27
1B3T	R	4,40	105,60	50,29	105,61	50,17
1MSE	S	0,06	1,23	0,61	2,21	0,85
1MSE	N	0,60	0,42	0,23	0,35	0,22
1MSE	E	-0,10	-41,68	-18,84	-43,14	-19,50
1MSE	S	0,10	2,42	1,52	2,22	1,25
1QRV	L	0,02	3,85	3,89	3,24	3,38
1QRV	M	1,20	8,78	8,67	8,50	8,30
1QRV	V	-0,30	3,59	3,92	3,87	4,32
1MNM	K	0,54	122,31	57,75	112,84	52,73
1MNM	E	0,24	-95,69	-43,63	-98,47	-44,81
1MNM	R	0,28	109,22	50,78	112,13	52,28
1MNM	R	0,81	143,26	74,12	144,14	74,77
1MNM	K	0,64	121,95	57,66	122,33	57,59
1MNM	I	0,88	0,98	1,55	0,98	1,54
1MNM	E	0,20	-99,02	-45,32	-93,44	-42,71
1MNM	I	0,86	1,97	1,20	1,60	0,99
1MNM	K	-0,13	84,02	38,40	85,38	39,07
1MNM	F	4,05	0,44	0,20	0,04	0,03
1MNM	I	4,05	1,21	0,77	1,27	0,79
1MNM	E	0,41	-75,20	-34,29	-76,24	-34,74
1MNM	N	1,75	-0,39	-0,13	-0,86	-0,28
1MNM	K	-0,13	77,57	35,31	80,69	36,83
1MNM	R	2,76	84,59	38,49	85,29	38,85
1MNM	H	-0,81	1,76	1,11	1,70	1,28
1MNM	V	0,58	-0,31	0,54	0,14	1,10
1MNM	T	0,64	-0,66	0,04	-0,58	0,08
1MNM	F	4,05	0,47	0,24	0,40	0,21
1MNM	S	0,43	-0,30	-0,06	-0,38	-0,10
1MNM	K	4,05	123,35	59,11	124,94	60,29
1MNM	R	4,05	139,10	67,04	139,33	66,83
1MNM	K	2,83	92,30	42,17	93,34	42,69
1MNM	H	-1,24	2,11	1,19	2,22	1,29
1MNM	I	3,24	1,49	0,73	1,36	0,67
1MNM	M	4,05	0,41	0,23	0,31	0,17
1MNM	K	4,05	125,53	58,61	126,36	58,99
1MNM	K	4,05	145,90	68,94	141,87	67,08
1MNM	F	-0,13	0,34	0,21	0,00	0,08
1MNM	E	2,27	-123,08	-56,10	-123,63	-56,24
1MNM	S	0	0,98	0,50	0,98	0,49
1MNM	V	0,11	2,10	0,89	1,76	0,76
1MNM	L	4,05	0,28	0,55	0,10	0,57
1MNM	T	4,05	1,54	0,73	1,15	0,57

1MNM	T	0,64	0,71	0,28	0,44	0,20
------	---	------	------	------	------	------

Apresenta-se de seguida (Figura III.2) a título de exemplo o gráfico de $\Delta\Delta G$ da mutação 2 do sistema 1BDT para 2 e 6 ns, $\epsilon=3$. Facilmente se observa que em ambas a dispersão de resultados é elevada e semelhante. No entanto o valor médio apresentado é semelhante: de 2,18 e 2,41 kcal mol⁻¹ para 2 e 6 ns, respetivamente, não existindo uma diferença significativa entre os valores de $\Delta\Delta G$ calculados para a simulação de 2 ns e para a simulação de 6 ns.

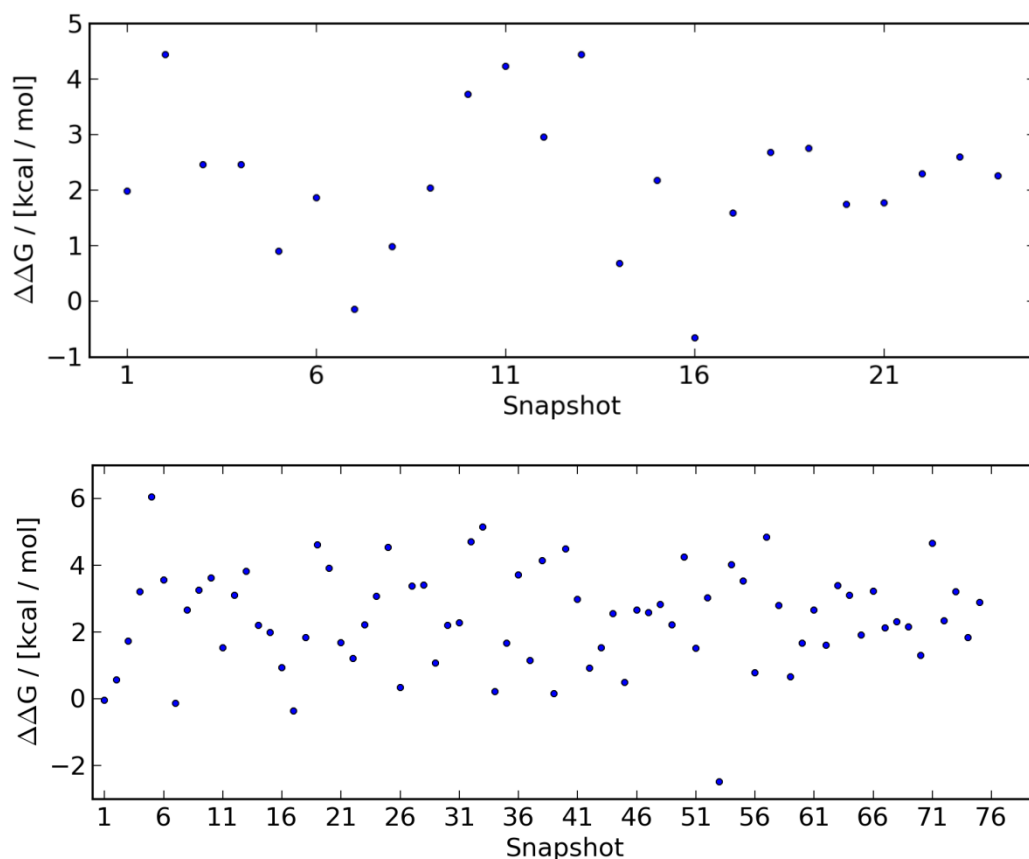


Figura III.2 - Representação da distribuição dos valores de $\Delta\Delta G_{MM-PBSA}$ (kcal mol⁻¹) para um conjunto de 25 (2 ns) e 75 estruturas (6 ns).

Apesar da simulação de um sistema proteína-DNA em 2 ns ser pequena, parece ser suficiente para o cálculo da energia livre. Este comportamento pode ser explicado pelo modo de ligação bem definido entre a proteína e o DNA, que implica rearranjos estruturais pouco significativos ao longo da simulação de DM. Não menos importante é o custo computacional associado a este tipo de cálculos. Por exemplo, para o complexo 1MNM, o tempo necessário para concluir o cálculo de cada mutação utilizando a simulação de DM de 6 ns foi cerca de 3,5 horas, por constante dielétrica, ou seja, cerca de 35 horas para realizar o cálculo de uma mutação com constantes dielétricas de 1 a 10. Para as 35 mutações e respetivas constantes dielétricas, cerca de 52 dias, para um só complexo. Por esta razão, e por se ter verificado a

Capítulo III: Resultados e Discussão

semelhança entre os valores, na simulação de 6 ns apenas se calculou com constante dielétrica de 1 a 7.

Calculou-se o erro associado às medições efectuadas com o nosso método teórico, tendo em conta os valores de energia livre de ligação experimentais (24), bem como a percentagem de resultados que obedecem a (25):

$$\text{Erro} = |\Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}}| \quad (24)$$

$$|\Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}}| < 1 \quad (25)$$

Os resultados obtidos podem ser consultados na Tabela III.2 e Tabela III.3.

Tabela III.2 – Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ para a simulação de 2 ns.

	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} / \text{kcal mol}^{-1}$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	114,78	129,55	40,55	31,30	25,74	22,05	19,40	17,42	29,80	14,64
Carregados	270,00	306,61	93,83	71,82	56,34	49,78	43,47	38,74	68,46	32,10
Polares	4,00	2,92	2,03	1,80	1,69	1,64	1,59	1,56	1,533	1,51
Apolares	3,81	3,25	2,99	2,93	2,90	2,87	2,86	2,85	2,87	2,84
	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} < 1 / \%$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	8	15	15	22	22	23	23	23	25	23
Carregados	--	--	--	--	--	--	--	--	--	--
Polares	17	39	39	50	50	50	50	50	56	50
Apolares	12	12	18	24	24	29	29	29	29	29

Tabela III.3 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ para a simulação de 6 ns.

	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} / \text{kcal mol}^{-1}$						
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$
Todos	115,11	65,68	40,66	31,37	25,80	22,52	19,44
Carregados	270,60	153,93	94,06	71,99	58,74	50,97	43,59
Polares	4,46	2,61	2,17	1,90	1,73	1,62	1,58
Apolares	3,61	2,67	2,89	2,84	2,83	2,81	2,83
	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} < 1 / \%$						
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$
Todos	10	13	15	20	22	27	23
Carregados	--	--	--	--	--	--	--
Polares	22	22	28	44	44	56	50
Apolares	12	24	24	24	29	35	29

Para as medições com a DM de 2 ns obtiveram-se erros de 40,55 e 19,40 kcal mol⁻¹ para as constantes dielétricas 3 e 7, respetivamente, enquanto para a DM de 6 ns obtiveram-se erros de 40,66 e 19,44 kcal mol⁻¹.

Relativamente aos valores obtidos de energia livre de ligação para as mutações estudadas é facilmente perceptível, tanto para a simulação de 2 ns como para a de 6 ns, que existe um erro

por excesso associado a todos os resíduos carregados (Arg, Lys, His, Asp, Glu). Na simulação de 2 ns verifica-se que os cálculos para os resíduos polares são mais concordantes com os valores experimentais do que os resíduos apolares, conseguindo-se ter uma razão $|\Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{exp}| < 1$ de 56 % ($\epsilon=9$) para os polares, enquanto que para os apolares este valor decresce para 29 % ($\epsilon=9$).

Como o objetivo é obter um método capaz de calcular a energia livre com menor erro experimental possível, usando o mínimo de tempo de computação, optar-se-ia pelo uso de 2 ns como tempo de simulação para os cálculos.

1.4. Importância do tipo de solvente na determinação de $\Delta\Delta G$

De seguida resolveu-se avaliar que influência o método de solvente utilizado na simulação de DM teria na determinação da energia livre de ligação do conjunto de resíduos em estudo. Para tal, os cálculos da energia livre de ligação foram efetuados para um conjunto de 25 estruturas geradas dos últimos 2 ns das simulações de DM efetuadas em solvente implícito para os cinco complexos proteína-DNA. Os resultados obtidos no cálculo em solvente implícito foram comparados com os resultados em solvente explícito de forma a atestar da influência do tipo de solvente na medição. Apenas os resultados referentes às constantes dielétricas 3 e 9 (Tabela III.4) serão apresentados. A totalidade dos resultados pode ser consultada no Capítulo VI - Informação de Suporte (Tabela VI.2).

Tabela III.4 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e em solvente implícito. Constantes dielétricas 3 e 9.

Proteína	AA Mutado	$\Delta\Delta G_{exp} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{MM-PBSA} / \text{kcal mol}^{-1}$			
			Solvente Explícito		Solvente Implícito	
			$\epsilon 3$	$\epsilon 9$	$\epsilon 3$	$\epsilon 9$
1BDT	M	1,10	4,68	4,87	1,38	1,29
1BDT	S	1,30	2,18	0,36	2,85	-0,52
1BDT	K	1,30	76,46	56,03	94,76	34,39
1BDT	M	1,90	-0,36	-0,21	0,69	0,31
1BDT	Q	1,80	4,42	3,69	4,01	4,38
1BDT	N	2,00	3,40	3,08	4,00	3,43
1BDT	R	6,00	112,64	83,29	124,66	49,31
1BDT	R	0,20	95,61	70,19	96,42	37,22
1BDT	N	-1,00	1,42	1,04	1,31	0,46
1BDT	R	-0,30	71,15	51,56	72,07	25,85
1BDT	S	-2,30	4,14	1,63	5,32	1,24

Capítulo III: Resultados e Discussão

1BDT	V	-2,30	3,71	1,48	4,01	1,78
1BDT	N	3,20	-1,08	-0,37	0,36	0,73
1BDT	S	-0,20	0,90	0,29	2,16	0,92
1BDT	Q	-0,40	0,08	-0,08	-0,57	-0,17
1B3T	R	3,41	110,52	84,56	118,15	90,42
1B3T	Y	2,62	8,08	5,82	5,51	6,11
1B3T	R	4,40	105,60	72,46	105,00	72,25
1MSE	S	0,06	1,23	0,60	4,85	1,73
1MSE	N	0,60	0,42	0,47	0,32	0,17
1MSE	E	-0,10	-41,68	-30,90	-44,18	-15,82
1MSE	S	0,10	2,42	1,35	4,02	1,66
1QRV	L	0,02	3,85	3,91	2,11	1,79
1QRV	M	1,20	8,78	8,66	5,33	5,96
1QRV	V	-0,30	3,59	3,88	5,73	3,11
1MNM	K	0,54	122,31	89,67	119,17	88,08
1MNM	E	0,24	-95,69	-71,11	-108,98	-78,01
1MNM	R	0,28	109,22	79,67	114,72	82,57
1MNM	R	0,81	143,26	108,44	140,05	105,14
1MNM	K	0,64	121,95	90,65	127,79	92,81
1MNM	I	0,88	0,98	1,62	1,55	1,60
1MNM	E	0,20	-99,02	-73,25	-85,62	-64,26
1MNM	I	0,86	1,97	0,94	0,99	0,25
1MNM	K	-0,13	84,02	62,71	85,67	64,06
1MNM	F	4,05	0,44	0,05	0,15	0,10
1MNM	I	4,05	1,21	0,74	1,10	0,73
1MNM	E	0,41	-75,20	-56,99	-76,43	-57,73
1MNM	N	1,75	-0,39	-0,61	0,29	-0,08
1MNM	K	-0,13	77,57	57,37	109,56	77,20
1MNM	R	2,76	84,59	62,91	89,47	66,31
1MNM	H	-0,81	1,76	0,85	3,36	2,48
1MNM	V	0,58	-0,31	0,63	1,16	3,06
1MNM	T	0,64	-0,66	-0,37	2,76	1,06
1MNM	F	4,05	0,47	0,09	0,11	0,03
1MNM	S	0,43	-0,30	0,23	-0,04	0,37
1MNM	K	4,05	123,35	93,45	137,19	97,83
1MNM	R	4,05	139,10	98,76	140,75	97,22
1MNM	K	2,83	92,30	68,20	94,59	69,64
1MNM	H	-1,24	2,11	0,68	1,06	-0,30
1MNM	I	3,24	1,49	0,47	1,27	0,21
1MNM	M	4,05	0,41	0,10	0,94	0,31
1MNM	K	4,05	125,53	91,34	134,64	94,65
1MNM	K	4,05	145,90	101,15	148,41	97,50
1MNM	F	-0,13	0,34	0,11	0,08	0,03
1MNM	E	2,27	-123,08	-89,04	-124,21	-88,92
1MNM	S	0	0,98	0,66	-0,11	-0,02
1MNM	V	0,11	2,10	0,71	1,77	0,72
1MNM	L	4,05	0,28	0,19	1,56	0,59

1MNM	T	4,05	1,54	0,75	1,39	0,80
1MNM	T	0,64	0,71	0,05	0,53	0,22

Tal como sucede nos cálculos em solvente explícito, também para os cálculos efetuados em solvente implícito existe um erro por excesso nos valores de energia livre de ligação para resíduos carregados (Tabela III.5). Este erro vai desde 282,99 kcal mol⁻¹ para $\epsilon=1$, até 33,84 kcal mol⁻¹ para $\epsilon=10$. O erro para a medição em solvente implícito é ligeiramente superior ao verificado para solvente explícito. Estes valores, claramente errados, têm a tendência a diminuir com o aumento da constante dielétrica interna utilizada, mas nunca se aproximam dos valores experimentais. Como o mesmo comportamento se verifica para os dois tipos de solvente e para dois tempos de simulação utilizados, leva-nos a concluir que o problema da não deteção de valores de energia livre de ligação corretos para os resíduos carregados reside na raiz do método, que não é apropriado para este tipo de sistemas proteína-DNA.

Tabela III.5 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{MM-PBSA}$ utilizando solvente implícito.

	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} / \text{kcal mol}^{-1}$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	120,37	120,63	42,41	32,71	26,90	23,03	20,26	18,18	28,09	15,47
Carregados	282,99	285,29	98,53	75,47	61,63	52,38	45,77	40,81	64,62	33,84
Polares	5,00	3,15	2,21	1,93	1,80	1,74	1,69	1,66	1,55	1,53
Apolares	3,37	2,88	2,45	2,42	2,42	2,42	2,41	2,41	2,46	2,41
	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} < 1 / \%$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	18	20	17	17	18	20	18	18	20	18
Carregados	--	--	--	--	--	--	--	--	4	--
Polares	28	39	28	33	33	39	33	33	33	33
Apolares	29	29	29	24	29	29	29	29	29	29

Relativamente a resíduos polares e apolares encontram-se algumas diferenças entre os resultados obtidos de solvente explícito e de implícito. Em solvente explícito, como vimos anteriormente, os valores de energia de 56% das mutações de resíduos polares ($\epsilon=9$) satisfazem a condição $|\Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{exp}| < 1$. Em solvente implícito esta percentagem desce para 39% ($\epsilon=2$ e $\epsilon=6$), o que significa que este modelo de solvente não é o mais indicado no tratamento deste tipo de resíduos. Para resíduos apolares a percentagem é igual, com 29% para solvente explícito e implícito. A explicação reside no facto da estrutura e dinâmica de sistemas proteína-DNA serem grandemente influenciados pela presença de iões, que não foram introduzidos no caso das DM em solvente implícito. A substituição de solvente explícito por solvente implícito oferece vantagens do ponto de vista computacional, tornando o cálculo mais rápido, mas por vezes a aproximação dos efeitos do meio não é suficiente para

Capítulo III: Resultados e Discussão

simular corretamente os rearranjos conformacionais dos sistemas moleculares. Para além disso as águas têm um papel importante ao nível das interfaces proteína-DNA, sendo fundamentais na interação com resíduos polares e carregados. A sua ausência resulta em valores de energia livre de ligação estimados com um erro superior quer para resíduos carregados quer para resíduos polares.

Da análise efetuada nos dois pontos anteriores pode-se concluir que o tempo de simulação não é um fator preponderante no cálculo da energia livre de ligação, ao contrário do que acontece com o tipo de solvente utilizado nas simulações de DM, pois obtêm-se melhores resultados em solvente explícito.

1.5. Influência do número de DM no cálculo de $\Delta\Delta G$

Nos pontos anteriores testou-se a influência do tempo de simulação e da representação do solvente no cálculo da energia livre de ligação. Depois disso resolveu-se avaliar a influência do número de DM utilizadas no respetivo cálculo, introduzindo para isso uma pequena alteração ao método. Nas situações anteriores utilizou-se uma DM, da estrutura WT do complexo, que foi pós-processada nos respetivos componentes (proteína + DNA) e que após tratamento deu origem ao complexo mutado que, também ele, foi separado nos seus componentes: proteína mutada e DNA. Neste novo teste, para além da DM do complexo WT, utiliza-se uma DM da estrutura WT da proteína, sendo ambas sujeitas a pós-processamento e mutagénesse. Um esquema exemplificativo das duas formas referidas anteriormente pode ser consultado na Figura III.3. Os resultados obtidos neste cálculo foram comparados com os resultados em solvente explícito. Da mesma forma, apenas os resultados referentes às constantes dielétricas 3 e 9 (Tabela III.6) serão apresentados. A totalidade dos resultados pode ser consultada no Capítulo VI - Informação de Suporte (Tabela VI.3).

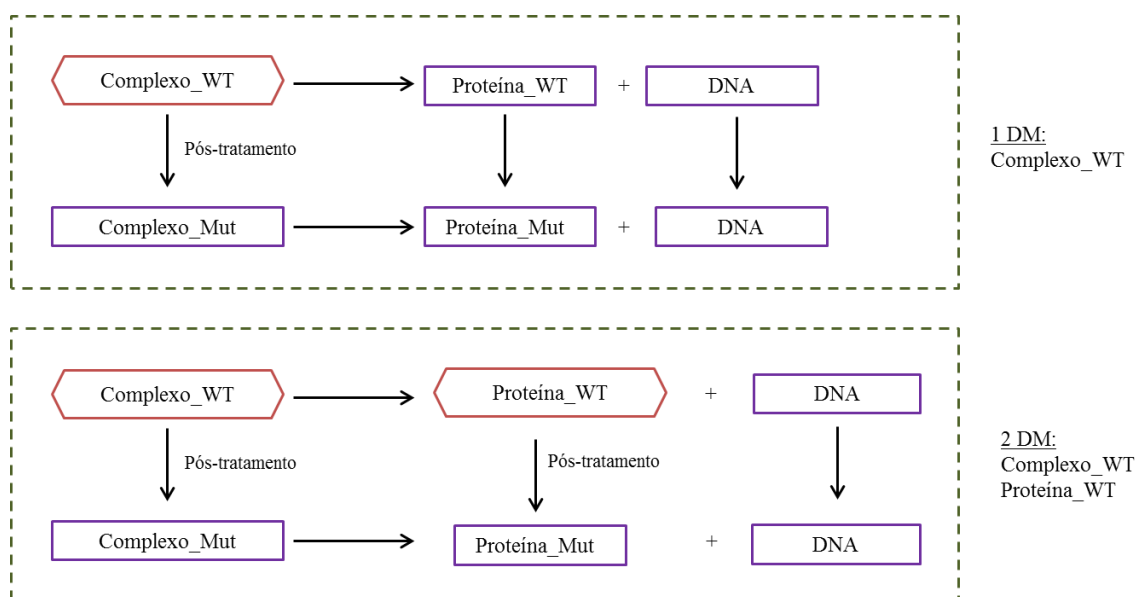


Figura III.3 - Representação esquemática das diferenças entre os dois métodos.

Tabela III.6 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e utilizando outra alternativa para o cálculo da energia livre de ligação, utilizando duas trajetórias. Constantes dielétricas 3 e 9.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$			
			2 ns		Monómeros	
			$\epsilon 3$	$\epsilon 9$	$\epsilon 3$	$\epsilon 9$
1BDT	M	1,10	4,68	4,87	4,10	5,30
1BDT	S	1,30	2,18	0,36	1,23	-0,24
1BDT	K	1,30	76,46	56,03	76,66	56,77
1BDT	M	1,90	-0,36	-0,21	0,23	-0,13
1BDT	Q	1,80	4,42	3,69	2,56	1,85
1BDT	N	2,00	3,40	3,08	2,35	1,69
1BDT	R	6,00	112,64	83,29	113,88	81,48
1BDT	R	0,20	95,61	70,19	94,55	69,39
1BDT	N	-1,00	1,42	1,04	1,88	1,40
1BDT	R	-0,30	71,15	51,56	70,60	50,68
1BDT	S	-2,30	4,14	1,63	2,62	0,63
1BDT	V	-2,30	3,71	1,48	3,01	1,06
1BDT	N	3,20	-1,08	-0,37	-0,33	0,16
1BDT	S	-0,20	0,90	0,29	0,55	0,65
1BDT	Q	-0,40	0,08	-0,08	-0,32	0,04
1B3T	R	3,41	110,52	84,56	109,49	82,71
1B3T	Y	2,62	8,08	5,82	9,21	7,62
1B3T	R	4,40	105,60	72,46	104,47	71,86
1MSE	S	0,06	1,23	0,60	0,40	0,49
1MSE	N	0,60	0,42	0,47	4,18	3,11
1MSE	E	-0,10	-41,68	-30,90	-42,92	-31,20

Capítulo III: Resultados e Discussão

1MSE	S	0,10	2,42	1,35	1,51	1,04
1QRV	L	0,02	3,85	3,91	3,22	3,73
1QRV	M	1,20	8,78	8,66	9,22	9,11
1QRV	V	-0,30	3,59	3,88	1,05	2,61
1MNM	K	0,54	122,31	89,67	119,08	87,01
1MNM	E	0,24	-95,69	-71,11	-93,16	-65,48
1MNM	R	0,28	109,22	79,67	106,61	77,34
1MNM	R	0,81	143,26	108,44	130,74	97,68
1MNM	K	0,64	121,95	90,65	118,98	86,91
1MNM	I	0,88	0,98	1,62	-3,70	-2,49
1MNM	E	0,20	-99,02	-73,25	-97,34	-69,70
1MNM	I	0,86	1,97	0,94	1,38	1,50
1MNM	K	-0,13	84,02	62,71	82,10	62,03
1MNM	F	4,05	0,44	0,05	0,23	0,00
1MNM	I	4,05	1,21	0,74	-0,23	-0,66
1MNM	E	0,41	-75,20	-56,99	-77,68	-56,75
1MNM	N	1,75	-0,39	-0,61	0,70	0,49
1MNM	K	-0,13	77,57	57,37	74,73	56,73
1MNM	R	2,76	84,59	62,91	80,56	59,04
1MNM	H	-0,81	1,76	0,85	-2,02	-2,57
1MNM	V	0,58	-0,31	0,63	0,73	1,32
1MNM	T	0,64	-0,66	-0,37	-0,89	-0,09
1MNM	F	4,05	0,47	0,09	0,66	-0,16
1MNM	S	0,43	-0,30	0,23	-0,56	-0,19
1MNM	K	4,05	123,35	93,45	120,21	91,52
1MNM	R	4,05	139,10	98,76	135,22	95,15
1MNM	K	2,83	92,30	68,20	90,77	67,58
1MNM	H	-1,24	2,11	0,68	1,51	0,66
1MNM	I	3,24	1,49	0,47	0,04	-0,81
1MNM	M	4,05	0,41	0,10	0,44	0,12
1MNM	K	4,05	125,53	91,34	122,58	88,70
1MNM	K	4,05	145,90	101,15	141,57	99,41
1MNM	F	-0,13	0,34	0,11	0,95	0,72
1MNM	E	2,27	-123,08	-89,04	-120,40	-86,48
1MNM	S	0	0,98	0,66	0,37	-0,16
1MNM	V	0,11	2,10	0,71	-0,27	-0,57
1MNM	L	4,05	0,28	0,19	-0,88	-1,25
1MNM	T	4,05	1,54	0,75	0,47	-0,58
1MNM	T	0,64	0,71	0,05	0,51	-0,62

Os resultados obtidos são semelhantes aos de solvente explícito, com o mesmo tipo de erro associado aos resíduos carregados. O erro médio para o conjunto de mutações analisadas varia entre 126,28 ($\epsilon=1$) e 13,84 ($\epsilon=10$) kcal mol⁻¹, sendo que o erro associado aos resíduos carregados varia entre 297,73 e 29,94 kcal mol⁻¹. Em relação aos resíduos polares e apolares verifica-se que existe uma ligeira diferença entre os valores obtidos neste cálculo e aqueles

obtidos utilizando a trajetória simples de 2 ns (solvente explícito), sendo que este método estima mais corretamente os resíduos polares e menos os resíduos apolares (Tabela III.7).

Não foi possível estudar na plenitude a influência do número de DM no cálculo de $\Delta\Delta G$ pois, até à data de escrita deste documento, não tínhamos disponíveis as DM do complexo mutado.

Devido ao custo computacional necessário para realizar novas simulações de DM para todos os complexos, contendo a mutação, optou-se por realizar simulações de DM dos 3 complexos mutados do 1B3T (1B3T:R9A, 1B3T:Y58A e 1B3T:R62A). Estes cálculos serão realizados posteriormente.

Tabela III.7 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{MM-PBSA}$ utilizando várias trajetórias.

	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} / \text{kcal mol}^{-1}$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	126,28	128,86	39,63	30,40	24,88	21,19	18,57	16,60	29,09	13,84
Carregados	297,73	304,46	91,70	69,62	56,38	47,56	41,27	36,54	66,38	29,94
Polares	4,20	3,10	1,83	1,70	1,64	1,60	1,58	1,57	1,62	1,55
Apolares	3,43	3,77	3,09	3,12	3,14	3,16	3,17	3,18	3,33	3,19
	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} < 1 / \%$									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	13	15	23	25	27	25	25	25	22	25
Carregados	--	4	--	--	--	--	--	--	--	--
Polares	28	28	56	56	67	61	61	39	50	61
Apolares	18	18	24	24	24	24	24	24	24	24

1.6. Importância da força iônica na determinação de $\Delta\Delta G$

Tal como foi discutido nas secções anteriores deste trabalho, os resultados obtidos para resíduos carregados estavam associados a um erro elevado. Este fato era ainda mais importante pois os resíduos carregados constituíam 40 % das mutações em estudo. Este problema era comum aos cálculos com trajetórias de 2 ns, 6 ns, solvente implícito e múltiplas trajetórias, o que significa que o problema estaria na raiz do método computacional utilizado. Efetivamente a forma linear da equação de PB, a qual é normalmente utilizada nos cálculos de ASM, não é a mais adequada para tratar sistemas altamente carregados, como é o caso de sistemas proteína-DNA, nos quais vigoram maioritariamente interações eletrostáticas e iônicas e cujo valor de energia livre de ligação é dependente da concentração iônica. Os ácidos nucleicos possuem uma densidade eletrônica negativa mais ou menos uniforme que resulta numa acumulação de contra-íões carregados positivamente em seu redor. A utilização da equação não-linear de PB para o tratamento de sistemas envolvendo DNA foi descrita anteriormente [87-89]. Deste modo, de forma a contabilizar a influência iônica decidiu-se

Capítulo III: Resultados e Discussão

aplicar a equação não-linear de PB, implementada no programa DelPhi, no cálculo da energia livre de ligação para o conjunto das mutações em estudo. Utilizaram-se as simulações de DM de 2 ns em solvente explícito e uma concentração de sal de 0,01M, que se encontra num intervalo fisiológico.

Tal como se fez nos pontos anteriores, os resultados obtidos foram comparados com os obtidos para a simulação de DM de 2 ns em solvente explícito, utilizando a equação linear de PB. Da mesma forma, apenas os resultados referentes às constantes dielétricas 3 e 9 (Tabela III.8) serão apresentados. A totalidade dos resultados pode ser consultada no Capítulo VI - Informação de Suporte (Tabela VI.4).

Tabela III.8 - Resultados dos cálculos de energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação linear e não-linear de Poisson-Boltzmann. Constantes dielétricas 3 e 9.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$			
			PB linear		PB não-linear	
			$\epsilon 3$	$\epsilon 9$	$\epsilon 3$	$\epsilon 9$
1BDT	M	1,10	4,68	4,87	4,60	4,93
1BDT	S	1,30	2,18	0,36	2,52	0,51
1BDT	K	1,30	76,46	56,03	2,42	5,13
1BDT	M	1,90	-0,36	-0,21	0,13	0,13
1BDT	Q	1,80	4,42	3,69	4,65	4,23
1BDT	N	2,00	3,40	3,08	3,28	3,23
1BDT	R	6,00	112,64	83,29	5,95	10,55
1BDT	R	0,20	95,61	70,19	2,41	6,62
1BDT	N	-1,00	1,42	1,04	-0,02	0,09
1BDT	R	-0,30	71,15	51,56	3,09	4,64
1BDT	S	-2,30	4,14	1,63	3,20	1,10
1BDT	V	-2,30	3,71	1,48	3,47	1,57
1BDT	N	3,20	-1,08	-0,37	-0,81	0,08
1BDT	S	-0,20	0,90	0,29	1,16	0,60
1BDT	Q	-0,40	0,08	-0,08	0,48	0,17
1B3T	R	3,41	110,52	84,56	6,65	17,51
1B3T	Y	2,62	8,08	5,82	8,51	7,11
1B3T	R	4,40	105,60	72,46	13,86	13,01
1MSE	S	0,06	1,23	0,60	1,36	0,56
1MSE	N	0,60	0,42	0,47	-0,45	-0,04
1MSE	E	-0,10	-41,68	-30,90	-0,20	-2,22
1MSE	S	0,10	2,42	1,35	2,65	1,46
1QRV	L	0,02	3,85	3,91	3,89	3,91
1QRV	M	1,20	8,78	8,66	9,02	8,74
1QRV	V	-0,30	3,59	3,88	3,51	3,95
1MNM	K	0,54	122,31	89,67	4,92	10,69
1MNM	E	0,24	-95,69	-71,11	2,05	-4,78

Capítulo III: Resultados e Discussão

1MNM	R	0,28	109,22	79,67	3,81	8,31
1MNM	R	0,81	143,26	108,44	13,45	23,46
1MNM	K	0,64	121,95	90,65	2,77	10,06
1MNM	I	0,88	0,98	1,62	0,48	1,46
1MNM	E	0,20	-99,02	-73,25	-0,78	-6,03
1MNM	I	0,86	1,97	0,94	1,39	0,83
1MNM	K	-0,13	84,02	62,71	-0,50	4,56
1MNM	F	4,05	0,44	0,05	0,11	0,05
1MNM	I	4,05	1,21	0,74	0,79	0,52
1MNM	E	0,41	-75,20	-56,99	1,38	-3,74
1MNM	N	1,75	-0,39	-0,61	0,63	0,24
1MNM	K	-0,13	77,57	57,37	1,65	4,72
1MNM	R	2,76	84,59	62,91	0,68	4,77
1MNM	H	-0,81	1,76	0,85	1,71	0,97
1MNM	V	0,58	-0,31	0,63	-0,30	0,67
1MNM	T	0,64	-0,66	-0,37	0,18	0,40
1MNM	F	4,05	0,47	0,09	0,42	0,17
1MNM	S	0,43	-0,30	0,23	-0,98	-0,23
1MNM	K	4,05	123,35	93,45	6,03	12,11
1MNM	R	4,05	139,10	98,76	16,69	17,78
1MNM	K	2,83	92,30	68,20	0,47	5,33
1MNM	H	-1,24	2,11	0,68	-1,53	0,99
1MNM	I	3,24	1,49	0,47	1,55	0,62
1MNM	M	4,05	0,41	0,10	0,56	0,22
1MNM	K	4,05	125,53	91,34	3,31	9,58
1MNM	K	4,05	145,90	101,15	7,21	14,44
1MNM	F	-0,13	0,34	0,11	0,17	0,10
1MNM	E	2,27	-123,08	-89,04	-3,02	-8,04
1MNM	S	0	0,98	0,66	0,26	0,20
1MNM	V	0,11	2,10	0,71	1,93	0,64
1MNM	L	4,05	0,28	0,19	-0,21	0,42
1MNM	T	4,05	1,54	0,75	0,67	0,33
1MNM	T	0,64	0,71	0,05	0,60	0,19

A principal diferença, e aquela que mais rapidamente se consegue observar, quando comparamos os resultados obtidos do cálculo utilizando a equação não-linear de PB e os anteriores utilizando a equação linear, é relativa ao comportamento dos resíduos carregados. Verificamos que já não existe uma determinação por excesso do valor de energia livre de ligação para estes resíduos mas sim valores próximos do esperado, quando se tem em consideração o erro inerente ao método utilizado. O erro teórico associado aos resíduos carregados varia entre 13,73 ($\epsilon=1$) e 7,20 ($\epsilon=10$) kcal mol⁻¹, quando anteriormente variava entre 270,0 ($\epsilon=1$) e 32,10 ($\epsilon=10$) kcal mol⁻¹. É uma diferença notória e que ganha maior relevância quando se compara o erro associado ao conjunto das sessenta mutações:

Capítulo III: Resultados e Discussão

anteriormente o erro variava entre 114,78 ($\epsilon=1$) e 14,64 ($\epsilon=10$) kcal mol⁻¹ e com a utilização da equação não-linear PB decresce para 7,92 ($\epsilon=1$) e 4,25 ($\epsilon=10$) kcal mol⁻¹. Os erros associados aos resíduos polares e apolares são, também eles, inferiores. Com a aplicação da equação não-linear foi possível identificar sete mutações (28%, $\epsilon=3$) cuja diferença para os valores experimentais é inferior a um; da mesma forma verificou-se um aumento de concordância para resíduos polares e apolares.

Desta forma torna-se evidente que a utilização da equação não-linear de PB é o método mais adequado para lidar com sistemas proteína-DNA.

Tabela III.9 - Resultados do erro associado ao cálculo de $\Delta\Delta G_{MM-PBSA}$ pela resolução da equação não-linear de PB.

	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} $ / kcal mol ⁻¹									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	7,92	3,22	2,76	3,19	3,54	3,79	4,00	4,09	4,19	4,25
Carregados	13,73	5,09	3,17	4,40	5,36	6,01	6,58	6,80	7,05	7,20
Polares	3,68	1,58	1,97	1,77	1,65	1,60	1,52	1,53	1,51	1,49
Apolares	3,84	2,20	2,99	2,90	2,87	2,85	2,83	2,83	2,83	2,83
	$ \Delta\Delta G_{MM-PBSA} - \Delta\Delta G_{Experimental} < 1$ / %									
	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
Todos	12	33	28	28	28	27	27	25	25	25
Carregados	4	16	28	21	8	4	0	0	0	0
Polares	22	50	33	44	56	56	61	56	56	56
Apolares	12	41	24	24	29	29	29	29	29	29

Seria interessante estudar o comportamento das mutações com a utilização de diferentes concentrações de força iónica. A título exemplificativo resolveu-se efetuar esse estudo para uma mutação do complexo 1BDT, utilizando uma concentração iónica de 0,1M. Os resultados obtidos (Tabela III.10) foram muito similares, o que nos impede de formular alguma conclusão.

Tabela III.10 - Comparação entre os valores obtidos para uma mutação utilizando concentrações iónicas de 0,1 e 0,01M.

C / M	AA	$\Delta\Delta G_{exp}$ / kcal mol ⁻¹	$\Delta\Delta G_{MM-PBSA}$ / kcal mol ⁻¹									
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
0.1	R	6,00	-7,63	1,66	5,76	7,75	8,96	9,74	10,30	10,70	11,01	11,24
0.01	R	6,00	-8,19	2,48	5,89	7,57	8,57	9,23	9,69	10,03	10,28	10,48

Durante a análise dos resultados observou-se uma tendência que nos despertou a atenção, relacionada com a contribuição relativa de cada termo para o cálculo da energia livre de ligação. Como vimos anteriormente, a energia livre de ligação é dada pela equação 23. Tal como se poderia esperar tendo em conta as características de sistemas proteína-DNA,

verificou-se que os termos que contribuíam de forma mais preponderante para a energia livre de ligação eram a soma dos termos referentes à energia eletrostática e à solvatação polar. Resolveu-se, portanto, estudar a existência de correlação entre os valores de $\Delta\Delta G_{\text{exp}}$ e $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$.

Os resultados podem ser consultados na Tabela III.11 e o respetivo tratamento de erros na Tabela III.12.

Por comparação da Tabela III.9 e Tabela III.12 conseguimos perceber que, globalmente, os erros obtidos são ligeiramente inferiores quando consideramos apenas a contribuição eletrostática.

Tabela III.11 - $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$ calculado pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação não-linear de Poisson-Boltzmann.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$										
			PB não-linear										
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$	
1BDT	M	1,10	-1,46	-0,66	-0,40	-0,28	-0,20	-0,15	-0,12	-0,09	-0,08	-0,06	
1BDT	S	1,30	9,13	4,68	3,20	2,45	2,00	1,70	1,48	1,32	1,19	1,09	
1BDT	K	1,30	-4,76	0,04	2,05	3,05	3,66	4,07	4,36	4,58	4,76	4,90	
1BDT	M	1,90	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	
1BDT	Q	1,80	1,79	0,90	0,60	0,45	0,36	0,29	0,25	0,21	0,19	0,16	
1BDT	N	2,00	0,32	0,19	0,15	0,12	0,11	0,10	0,10	0,09	0,09	0,09	
1BDT	R	6,00	-11,05	-0,26	3,23	4,98	6,03	6,72	7,20	7,55	7,83	8,04	
1BDT	R	0,20	-10,56	-0,60	2,69	4,31	5,27	5,89	6,33	6,66	6,90	7,09	
1BDT	N	-1,00	-0,33	-0,11	-0,03	0,01	0,03	0,05	0,06	0,07	0,08	0,08	
1BDT	R	-0,30	-1,33	1,89	2,98	3,54	3,88	4,12	4,29	4,42	4,53	4,62	
1BDT	S	-2,30	9,60	4,91	3,34	2,56	2,09	1,78	1,55	1,38	1,25	1,14	
1BDT	V	-2,30	8,57	4,32	2,89	2,18	1,75	1,47	1,26	1,11	0,99	0,89	
1BDT	N	3,20	-4,00	-2,00	-1,33	-1,00	-0,80	-0,67	-0,57	-0,50	-0,45	-0,40	
1BDT	S	-0,20	2,54	1,24	0,81	0,59	0,47	0,38	0,33	0,28	0,25	0,22	
1BDT	Q	-0,40	1,37	0,69	0,46	0,35	0,28	0,24	0,20	0,18	0,16	0,14	
1B3T	R	3,41	-43,81	-12,95	-3,25	1,24	3,75	5,30	6,34	7,08	7,61	8,01	
1B3T	Y	2,62	6,06	3,58	2,68	2,20	1,89	1,68	1,51	1,38	1,27	1,18	
1B3T	R	4,40	12,81	11,90	11,86	11,72	11,56	5,30	11,26	11,13	11,01	10,90	
1MSE	S	0,06	3,71	1,95	1,36	1,06	0,88	0,76	0,67	0,61	0,56	0,51	
1MSE	N	0,60	-1,74	-0,78	-0,46	-0,31	-0,22	-0,15	-0,11	-0,08	-0,06	-0,04	
1MSE	E	-0,10	-10,02	-6,50	-5,33	-4,75	-4,40	0,76	-4,02	-3,90	-3,81	-3,75	
1MSE	S	0,10	5,12	2,82	1,95	1,51	1,24	1,06	0,93	0,83	0,75	0,69	
1QRV	L	0,02	-0,03	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03	-0,03	-0,02	-0,02	
1QRV	M	1,20	1,49	0,79	0,55	0,43	0,36	0,31	0,28	0,25	0,23	0,21	
1QRV	V	-0,30	-2,03	-1,06	-0,73	-0,57	-0,47	-0,40	-0,36	-0,32	-0,29	-0,27	
1MNM	K	0,54	-15,51	-2,26	2,15	4,35	5,65	6,51	7,12	7,58	7,93	8,20	
1MNM	E	0,24	22,02	6,73	1,62	-0,94	-2,48	-3,51	-4,24	-4,79	-5,22	-5,56	
1MNM	R	0,28	-10,88	-0,69	2,72	4,42	5,43	6,10	6,58	6,94	7,21	7,43	

Capítulo III: Resultados e Discussão

1MNM	R	0,81	-35,48	-9,43	-1,09	2,90	5,19	6,65	7,65	8,37	8,91	9,32
1MNM	K	0,64	-21,97	-5,13	0,47	3,25	4,90	5,98	6,75	7,32	7,76	8,10
1MNM	I	0,88	-4,27	-2,11	-1,38	-1,01	-0,79	-0,64	-0,53	-0,46	-0,39	-0,35
1MNM	E	0,20	14,95	3,08	-0,86	-2,83	-4,01	-4,80	-5,36	-5,78	-6,11	-6,38
1MNM	I	0,86	2,41	1,16	0,87	0,66	0,53	0,45	0,39	0,34	0,31	0,28
1MNM	K	-0,13	-15,54	-4,29	-0,53	1,36	2,50	3,25	3,80	4,21	4,53	4,79
1MNM	F	4,05	0,25	0,12	0,09	0,07	0,05	0,04	0,04	0,03	0,03	0,02
1MNM	I	4,05	1,31	0,63	0,42	0,31	0,25	0,21	0,18	0,16	0,15	0,14
1MNM	E	0,41	16,47	5,15	1,37	-0,53	-1,68	-2,45	-5,06	-3,42	-3,75	-4,02
1MNM	N	1,75	1,60	0,76	0,48	0,33	0,24	0,18	0,96	0,11	0,08	0,06
1MNM	K	-0,13	-7,32	-0,66	1,58	2,71	3,40	3,86	4,19	4,45	4,65	4,81
1MNM	R	2,76	-10,95	-2,30	0,63	2,12	3,03	3,65	4,10	4,44	4,72	4,94
1MNM	H	-0,81	3,33	1,51	0,98	0,72	0,57	0,46	0,39	0,34	0,29	0,26
1MNM	V	0,58	-4,27	-0,07	-1,34	-0,99	-0,77	-0,63	-0,53	-0,46	-0,40	-0,35
1MNM	T	0,64	-1,03	-0,57	-0,50	-0,46	-0,42	-0,40	-0,38	-0,36	-0,34	-0,33
1MNM	F	4,05	1,11	0,50	0,33	0,25	0,20	0,17	0,14	0,12	0,11	0,10
1MNM	S	0,43	-3,27	-1,67	-1,07	-0,77	-0,59	-0,47	-0,39	-0,32	-0,28	-0,24
1MNM	K	4,05	-15,96	-1,78	2,85	5,20	6,59	7,50	8,13	8,60	8,96	9,24
1MNM	R	4,05	-1,30	10,73	12,60	13,30	13,60	13,73	13,76	13,75	13,71	12,26
1MNM	K	2,83	-13,80	-3,11	0,45	2,24	3,32	4,05	5,32	4,98	5,29	5,54
1MNM	H	-1,24	5,65	2,73	-3,71	1,34	1,06	0,88	0,74	0,64	0,57	0,50
1MNM	I	3,24	4,18	2,12	1,42	1,08	0,87	0,72	0,62	0,54	0,48	0,43
1MNM	M	4,05	1,50	0,77	0,53	0,40	0,33	0,27	0,24	0,21	0,19	0,17
1MNM	K	4,05	-17,10	-2,51	2,27	4,72	6,14	7,07	7,73	8,22	8,59	8,88
1MNM	K	4,05	-23,44	-1,98	4,41	7,36	9,00	10,00	10,67	11,13	11,46	11,70
1MNM	F	-0,13	0,31	0,19	0,13	0,10	0,08	0,07	0,06	0,06	0,05	0,05
1MNM	E	2,27	11,99	-0,43	-4,25	-6,12	-7,22	-7,94	-8,44	-8,81	-9,09	-9,31
1MNM	S	0	0,30	0,16	0,18	0,19	0,19	0,19	0,19	0,19	0,18	0,18
1MNM	V	0,11	5,80	2,90	1,92	1,44	1,14	0,95	0,81	0,71	0,63	0,56
1MNM	L	4,05	-2,80	-1,41	-0,89	0,03	0,17	-0,39	-0,32	-0,26	0,42	0,45
1MNM	T	4,05	1,58	0,78	0,55	0,44	0,37	0,32	0,29	0,26	0,24	0,22
1MNM	T	0,64	1,83	1,00	0,65	0,48	0,38	0,31	0,26	0,22	0,20	0,17

Tabela III.12 - Resultados do erro associado ao cálculo de $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$ pela resolução da equação não-linear de PB.

	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} / \text{kcal mol}^{-1}$									
	ε1	ε2	ε3	ε4	ε5	ε6	ε7	ε8	ε9	ε10
Todos	8,33	3,37	2,25	2,30	2,52	2,61	2,97	3,06	3,14	3,19
Carregados	15,47	4,98	2,70	3,00	3,62	3,88	4,80	5,01	5,26	5,39
Polares	3,21	2,07	1,66	1,52	1,45	1,40	1,32	1,33	1,33	1,32
Apolares	3,24	2,32	2,22	2,09	2,04	2,04	2,02	2,00	1,95	1,94
	$ \Delta\Delta G_{\text{MM-PBSA}} - \Delta\Delta G_{\text{Experimental}} < 1 / \%$									
	ε1	ε2	ε3	ε4	ε5	ε6	ε7	ε8	ε9	ε10
Todos	15	27	30	30	35	37	30	30	30	32
Carregados	4	12	24	20	16	16	0	0	0	0
Polares	22	39	39	44	61	67	67	61	61	61
Apolares	24	35	29	29	35	35	35	41	41	41

1.7. Importância da constante dielétrica na determinação de $\Delta\Delta G$

Finalmente decidiu-se avaliar a influência da constante dielétrica na determinação da energia livre de ligação, tendo em conta os resultados dos cálculos obtidos anteriormente. A utilização de múltiplas constantes dielétricas para diferentes grupos de resíduos foi extensivamente estudado no passado, com boas taxas de sucesso, para interfaces proteína-proteína. Como tal, foi definido que para resíduos polares, apolares e carregados se deveria utilizar constantes dielétricas de 3, 2 e 4, respetivamente. Para interfaces proteína-DNA o estudo é muito diminuto, havendo apenas alguns relatos de utilização de constantes dielétricas muito elevadas (maiores do que 50) [63]. Para efeitos de comparação utilizaram-se os dois resultados obtidos com a resolução da equação não-linear de PB, ou seja, $\Delta\Delta G_{MM-PBSA}$ e $\Delta\Delta E_{ele} + \Delta\Delta G_{solv\ polar}$ (Tabela III.9 e Tabela III.12). Adicionalmente o desempenho do método e das várias constantes será avaliado por um conjunto de testes estatísticos, nomeadamente:

$$F1 = \frac{2PR}{P+R} \quad (26)$$

$$P = VPP = \frac{VP}{VP+FP} \quad (27)$$

$$R = \text{Sensibilidade} = \frac{VP}{VP+FN} \quad (28)$$

$$E = \text{Especificidade} = \frac{VN}{FP+VN} \quad (29)$$

$$A = \text{Precisão} = \frac{VP+VN}{VP+FP+VN+FN} \quad (30)$$

O teste F1 é definido em função do Valor Preditivo Positivo (VPP) e da Sensibilidade, em que VP refere-se aos Verdadeiros Positivos (HS calculados que são efetivamente HS); FP refere-se aos Falsos Positivos (HS calculados que não são efetivamente HS); FN refere-se aos Falsos Negativos (HS não calculados que não são efetivamente HS) e VN refere-se aos Verdadeiros Negativos (NS corretamente calculados). A precisão é definida como o rácio entre o número de resíduos corretamente previstos (VP + VN) e o número total de resíduos previstos. A sensibilidade e a especificidade são importantes pois aferem do poder de previsão de HS e NS, respetivamente; o VPP avalia a precisão das previsões positivas enquanto os restantes dão uma ideia global da performance do método. Cada um dos valores obtidos foi classificado como VP, FP, VN ou FN e as equações descritas anteriormente foram aplicadas. Os resultados globais dos testes, para o conjunto das 60 mutações e com divisão por grupos carregados, polares e apolares, pode ser consultado na Tabela VI.5, Tabela VI.6, Tabela VI.7 e Tabela VI.8. Se se estivesse unicamente a analisar os resultados globais para o conjunto das sessenta mutações poder-se-ia dizer que $\epsilon=6$, utilizando apenas a contribuição da energia

Capítulo III: Resultados e Discussão

eletrostática e solvatação polar, seria a mais adequada para o sistema em causa tendo em conta os resultados do erro e dos vários testes estatísticos. Contudo, é preferível decompor o sistema em resíduos carregados, polares e apolares e procurar a melhor correspondência para cada um dos casos. Assim sendo, e após se avaliar os erros e os resultados dos testes estatísticos para os dois tipos de determinação da energia livre de ligação, concluiu-se que para os resíduos carregados a melhor concordância entre o erro médio e a correta determinação de HS e NS era obtida pela utilização de uma constante dielétrica 5, pelo método $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$. A formação de complexos proteína-DNA remove os resíduos carregados do solvente em redor enquanto leva a reorientação do sal em volta do DNA. Os efeitos eletrostáticos são portanto elevados em torno destes resíduos. Desta forma parece que tendo em conta só esta contribuição para a energia livre se obtém uma maior concordância com os valores experimentais. Para resíduos polares e apolares a melhor concordância foi obtida utilizando o cálculo de $\Delta\Delta G_{\text{MM-PBSA}}$ e uma constante dielétrica de 6 e 2, respetivamente.

Desta forma, depois de decidido qual a melhor constante dielétrica para cada um dos tipos de resíduo, resolveu-se avaliar a capacidade da metodologia computacional desenvolvida para condições que se aproximassem das “ideais”. Após se agruparem as sessenta mutações com os respetivos valores determinados computacionalmente, segundo as condições dielétricas descritas anteriormente (Tabela III.13) obteve-se um erro médio entre os valores teóricos e os valores experimentais de $2,86 \text{ kcal mol}^{-1}$ e valores de 43%, 60%, 50%, 60% e 60% para o VPP, Sensibilidade, F1, Precisão e Especificidade, respetivamente.

Tabela III.13 - Resultados globais em função do melhor método e constante dielétrica, para os grupos de resíduos carregados, polares e apolares.

AA Mutado	Mutação	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	Média / kcal mol^{-1}	Método
E	2	0,24	-2,48	$\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$ $\epsilon = 5$
E	3	-0,10	-4,40	
E	7	0,20	-4,01	
E	12	0,41	-1,68	
E	31	2,27	-7,22	
H	16	-0,81	0,57	
H	24	-1,24	1,06	
K	1	0,54	5,65	
K	3	1,30	3,66	
K	5	0,64	4,90	
K	9	-0,13	2,50	
K	14	-0,13	3,40	
K	21	4,05	6,59	

Capítulo III: Resultados e Discussão

K	23	2,83	3,32	
K	28	4,05	6,14	
K	29	4,05	9,00	
R	0	3,41	3,75	
R	2	4,40	11,56	
R	3	0,28	5,43	
R	4	0,81	5,19	
R	7	6,00	6,03	
R	8	0,20	5,27	
R	10	-0,30	3,88	
R	15	2,76	3,03	
R	22	4,05	13,60	
<hr/>				
F	10	4,05	0,16	
F	19	4,05	0,61	
F	30	-0,13	0,22	
I	6	0,88	-0,25	
I	8	0,86	1,83	
I	11	4,05	1,01	
I	26	3,24	2,24	
L	0	0,02	3,89	$\Delta\Delta G_{\text{MM-PBSA}}$
L	34	4,05	-0,70	
M	1	1,20	9,30	$\varepsilon = 2$
M	1	1,10	4,35	
M	4	1,90	0,13	
M	27	4,05	0,81	
V	2	-0,30	3,18	
V	12	-2,30	4,89	
V	17	0,58	0,64	
V	33	0,11	2,90	
<hr/>				
N	1	0,60	-0,14	
N	6	2,00	3,24	
N	9	-1,00	0,06	
N	13	3,20	-0,14	
N	13	1,75	0,34	
Q	5	1,80	4,34	
Q	15	-0,40	0,25	
S	0	0,06	0,77	$\Delta\Delta G_{\text{MM-PBSA}}$
S	4	0,10	1,76	
S	2	1,30	1,02	$\varepsilon = 6$
S	11	-2,30	1,63	
S	14	-0,20	0,73	
S	20	0,43	-0,42	
S	32	0,00	0,23	
T	18	0,64	0,33	
T	35	4,05	0,42	
T	36	0,64	0,29	
Y	1	2,62	7,51	

Capítulo III: Resultados e Discussão

Tabela III.14 - Resultado do erro médio e dos diferentes testes estatísticos para os resultados obtidos por aplicação dos métodos referidos anteriormente.

Erro médio / kcal mol⁻¹	VP	FP	FN	VN	P	R	F1	A	E
2,86	12	16	8	24	43	60	50	60	60

Estes valores, apesar de inferiores comparados com os obtidos para sistemas proteína-proteína usando o nosso método anteriormente descrito, encontram-se dentro da ordem de grandeza da maioria dos métodos desenvolvidos para proteína-proteína[90]. Salienta-se a escassa existência de trabalhos semelhantes para este tipo de interface na literatura.

O método computacional *Alanine Scanning Mutagenesis*, otimizado para complexos proteína-DNA neste trabalho, apresenta algumas diferenças relativamente ao anteriormente descrito para complexos proteína-proteína. Inicialmente obtém-se a estrutura *wild-type* do complexo que é sujeita a uma simulação de DM (preferencialmente em solvente explícito em oposição ao solvente implícito usado anteriormente), a um pós-tratamento em que ocorre a divisão nos monómeros e respectiva mutação. Aqui reside a principal diferença entre os dois métodos, uma vez que se utiliza a equação não linear de Poisson-Boltzmann de forma a contabilizar a ação da força iónica. Os resíduos apolares, polares e carregados devem ser calculados com constantes dielétricas de 2, 6 e 5, respetivamente, sendo que para os resíduos carregados apenas se deve contabilizar as componentes eletrostáticas e de solvatação polar.

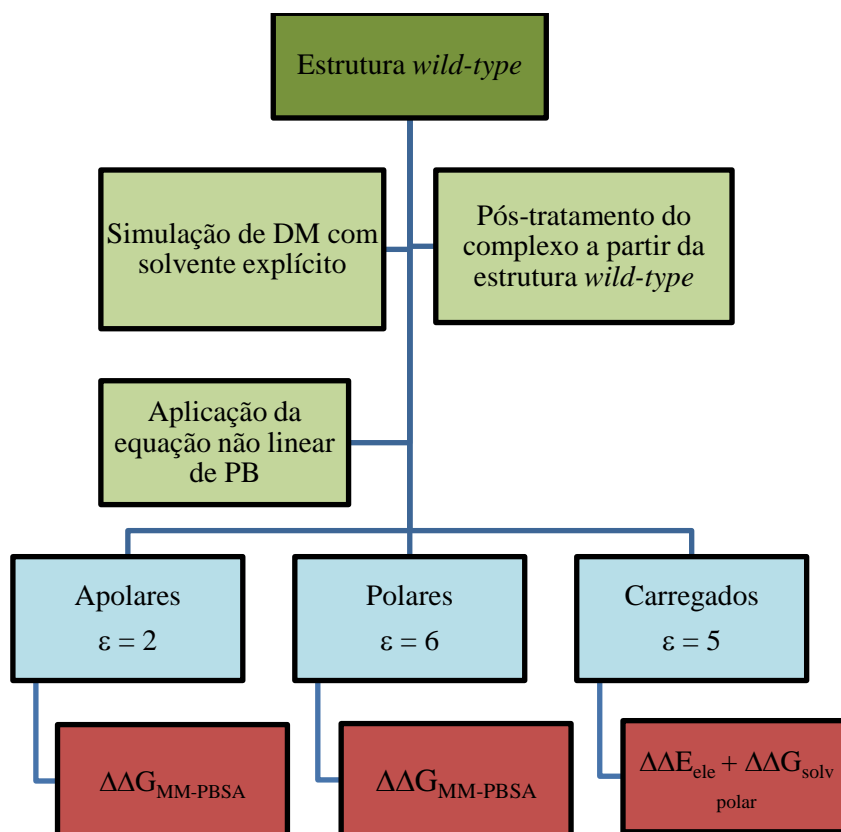


Figura III.4 - Representação esquemática da metodologia computacional ASM proposta para interfaces proteína-DNA.

1.8. RDFs

Apesar dos HS se encontrarem protegidos da ação da água, não nos podemos esquecer que a água tem um papel fundamental na determinação da estrutura e dinâmica da maioria dos sistemas biológicos, pela formação de interações específicas, mediação da transferência de prótons, sendo até utilizada como um componente estrutural na estrutura secundária das proteínas. Pretendíamos estudar o comportamento dinâmico das moléculas de água localizadas à volta da interface proteína-DNA nas simulações de DM efectuadas. As simulações de DM permitem-nos estudar a influência das moléculas de água nos HS e NS.

Geralmente, o RDF da água apresenta um perfil oscilatório e um pico devido à presença de ligações de hidrogénio. Mediram-se os RDFs para os 60 resíduos dos vários complexos analisados neste trabalho. A título exemplificativo, na Figura III.5 apresentam-se dois perfis de RDF, representando um HS e um NS. Como pode ser facilmente observado, o comportamento exibido por cada um deles é bastante diferente. O pico localizado aproximadamente a 3Å na Figura III.5-b, aparece devido à forte interação entre os átomos de hidrogénio da água e os átomos de oxigénio do grupo carbonilo do NS. Os picos seguintes,

Capítulo III: Resultados e Discussão

cuja localização varia de acordo com o sistema considerado, devem-se a interações entre as moléculas de água e os átomos (com a excepção do hidrogénio) do aminoácido. Estes picos são muito menos definidos no caso de um HS (Figura III.5-a). Os painéis b e d mostram que o número típico de águas à volta de um HS e de um NS são diferentes.

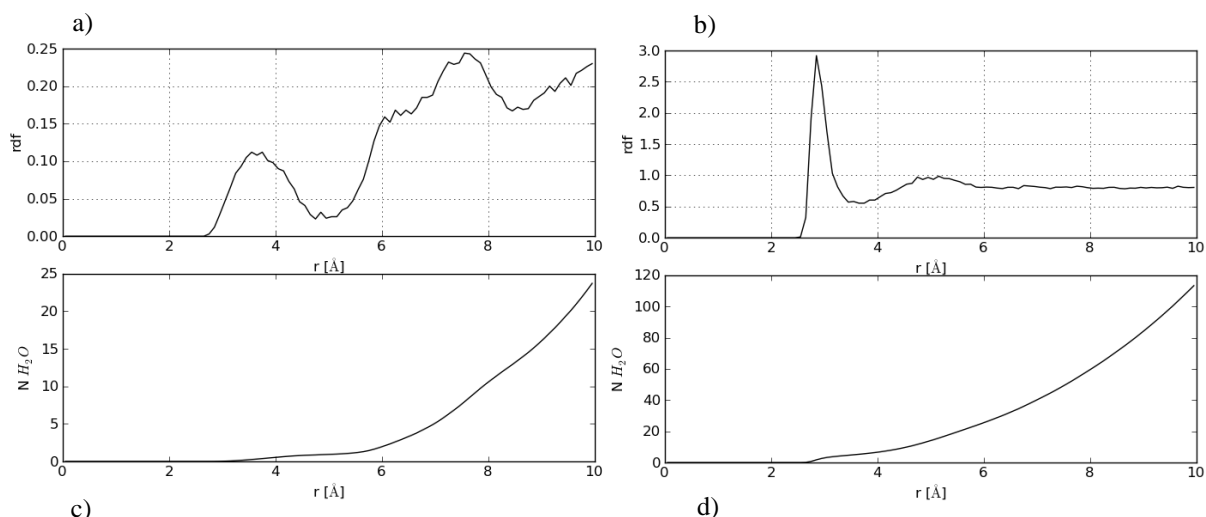


Figura III.5 - Representação do perfil de RDF e número médio de águas à volta de um HS (a, c) e de um NS (b, d).

Partindo dos cálculos de RDF para os 60 resíduos analisados das cinco interfaces proteína-DNA estudadas, mediram-se o número médio de moléculas de água localizadas à volta de cada um destes resíduos. A Tabela III.15 sumariza os resultados obtidos, em função de HS e NS, para o conjunto de resíduos analisados de cada um dos sistemas em estudo. A Tabela III.16 apresenta os resultados gerais para HS e NS. Estes valores foram medidos com um *cutoff* de 3, 4, 5 e 6 Å. Para um melhor entendimento dos resultados obtidos, estes valores foram representados graficamente (Figura III.6). Analisando estes dados torna-se perceptível o facto do número médio de águas em redor dos HS ser notoriamente inferior em comparação com os NS. Como seria expectável, um aumento do *cutoff* utilizado na medição resulta num aumento de moléculas de água localizadas em redor de um resíduo. Analisando os resultados gerais obtidos para HS e NS, que nos permite fazer uma melhor comparação entre os dois comportamentos, pois nem todos os sistemas eram constituídos simultaneamente por HS e NS, verificamos que, por exemplo, para uma distância de 5 Å existem cerca de oito moléculas de água em redor de um NS, enquanto para um HS existem apenas cinco. Para um NS, o número de águas que circunda um resíduo pode variar entre 1,19 e 14,16, em média.

A teoria do O-ring, formulada inicialmente por Thorn e Bogan [6] sugere que os HS encontram-se geralmente cercados por resíduos não importantes para a ligação, e cujo papel seria abrigá-los da ação do solvente. Baseado nesta teoria surgiu o conceito de que os HS teriam valores baixos de SASA. Os resultados aqui apresentados, que têm em conta a dinâmica de um sistema proteína-DNA, mostram claramente que o número médio de águas em redor de um HS é muito inferior em comparação com um NS. Este estudo, juntamente com outros realizados em interfaces proteína-proteína, confirmam que os HS encontram-se protegidos do solvente, tal como previsto pela teoria do O-ring.

Tabela III.15 - Número médio de moléculas de água localizadas a distâncias compreendidas entre 3 e 6 Å dos HS e NS, para cada um dos complexos estudados.

Complexo	1B3T		1BDT		1MNM		1MSE		1QRV	
Distância / Å	HS	NS	HS	NS	HS	NS	HS	NS	HS	NS
3	0,21	--	0,84	0,89	1,33	1,41	--	1,79	--	0,01
4	1,87	--	2,23	3,12	3,13	4,20	--	4,23	--	1,28
5	3,98	--	3,66	6,50	6,19	9,45	--	8,97	--	3,24
6	5,93	--	5,90	11,22	10,68	16,87	--	14,76	--	6,16

Tabela III.16 - Número médio de águas localizadas a distâncias compreendidas entre 3 e 6 Å para o conjunto de HS e NS em estudo.

Distância / Å	HS	NS
3	0,93	1,19
4	2,59	3,66
5	5,01	8,05
6	8,30	14,16

Capítulo III: Resultados e Discussão

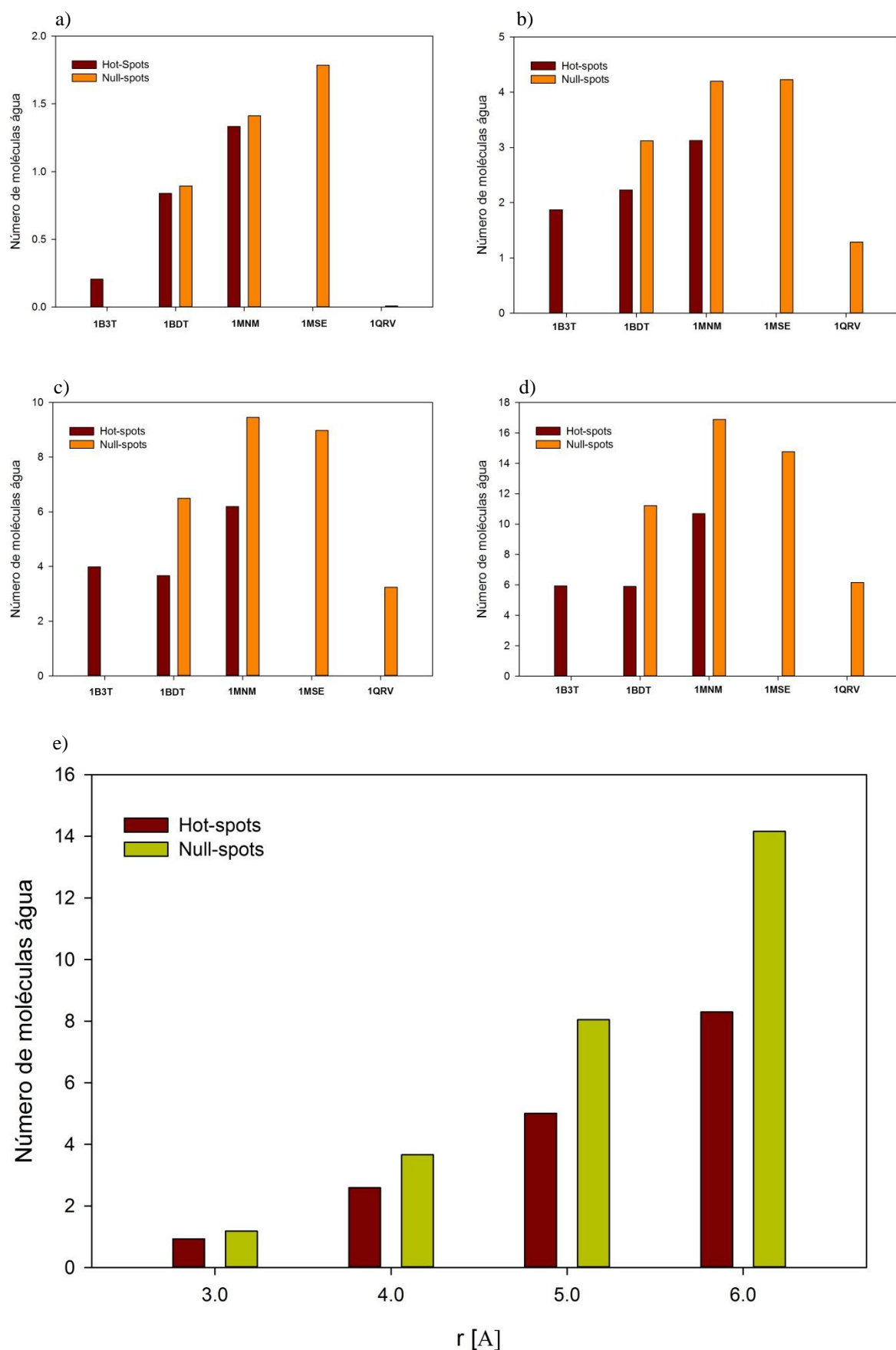


Figura III.6 - Representação dos valores médios de moléculas de água em redor de HS e NS para distâncias de 3 Å (a), 4 Å (b), 5 Å (c) e 6 Å (d), para os cinco complexos estudados.

1.9. SASAs

Diferentes cálculos de SASA foram efetuados de forma a se completar o estudo da importância de moléculas de água no microambiente do HS e NS. O SASA de cada resíduo, pertencentes à base de dados dos cinco complexos analisados, foi medido nos últimos 2 ns da DM em solvente explícito. Δ SASA e relSASA, definidos pelas equações 25 e 26, respetivamente, foram também medidos. relSASA permite diferenciar resíduos com igual Δ SASA mas diferente exposição ao solvente, como por exemplo, um resíduo com 50 Å² de acessibilidade ao solvente no monómero e 0 Å² no complexo de um resíduo com 150 Å² de acessibilidade ao solvente no monómero e 100 Å² no complexo. Em ambos os casos Δ SASA é 50 Å² mas a importância da acessibilidade ao solvente é diferente.

$$\Delta SASA = SASA_{cpx} - SASA_{mon} \quad (25)$$

$$relSASA = \frac{SASA_{cpx} - SASA_{mon}}{SASA_{mon}} = \frac{\Delta SASA}{SASA_{mon}} \quad (26)$$

Estudos anteriores realizados por Thor e Bogan [6], Guharoy *et al.*[91], Moreira *et al.* [92], entre outros autores, em interfaces proteína-proteína referem o importante papel do solvente na determinação da contribuição energética de um determinado resíduo para a ligação, bem como a correlação entre HS e valores baixos de SASA, uma vez que os HS estariam protegidos da ação do solvente por outros resíduos.

Os nossos resultados mostram que elevados valores de Δ SASA e relSASA são condição necessária para um resíduo ser considerado HS. Contudo, existem NS que também apresentam valores elevados de Δ SASA, e como tal esta característica não é suficiente para fazer a distinção entre HS e NS. Uma vez que propriedades como a hidrofobicidade / hidrofiliidade são fatores importantes na interface, decidiu-se avaliar o efeito da natureza do resíduo na relação entre o SASA e a energia livre de ligação. Na Figura III.7 estão representados os valores médios e desvios (Δ SASA e relSASA) para todos os resíduos estudados, resíduos carregados (Asp, Glu, His, Lys e Arg), polares (Ser, Thr, Asn, Gln e Tyr), apolares (Val, Ile, Leu, Met, Phe e Trp) e aromáticos (Phe, Trp, Tyr e His). Como pode ser observado, existe uma aparente diferença entre os valores médios de Δ SASA para HS e NS, com exceção do conjunto total de resíduos em que a diferença de HS para NS é mínima, sendo contudo o valor inferior para os HS. Tal deve-se ao facto de não existirem praticamente resíduos apolares como HS, o que influencia esta medida estatística. Relativamente à

Capítulo III: Resultados e Discussão

contribuição de cada tipo de resíduo, verifica-se que existe uma clara distinção para os carregados e aromáticos, em que o valor para os NS é inferior a metade do valor para HS.

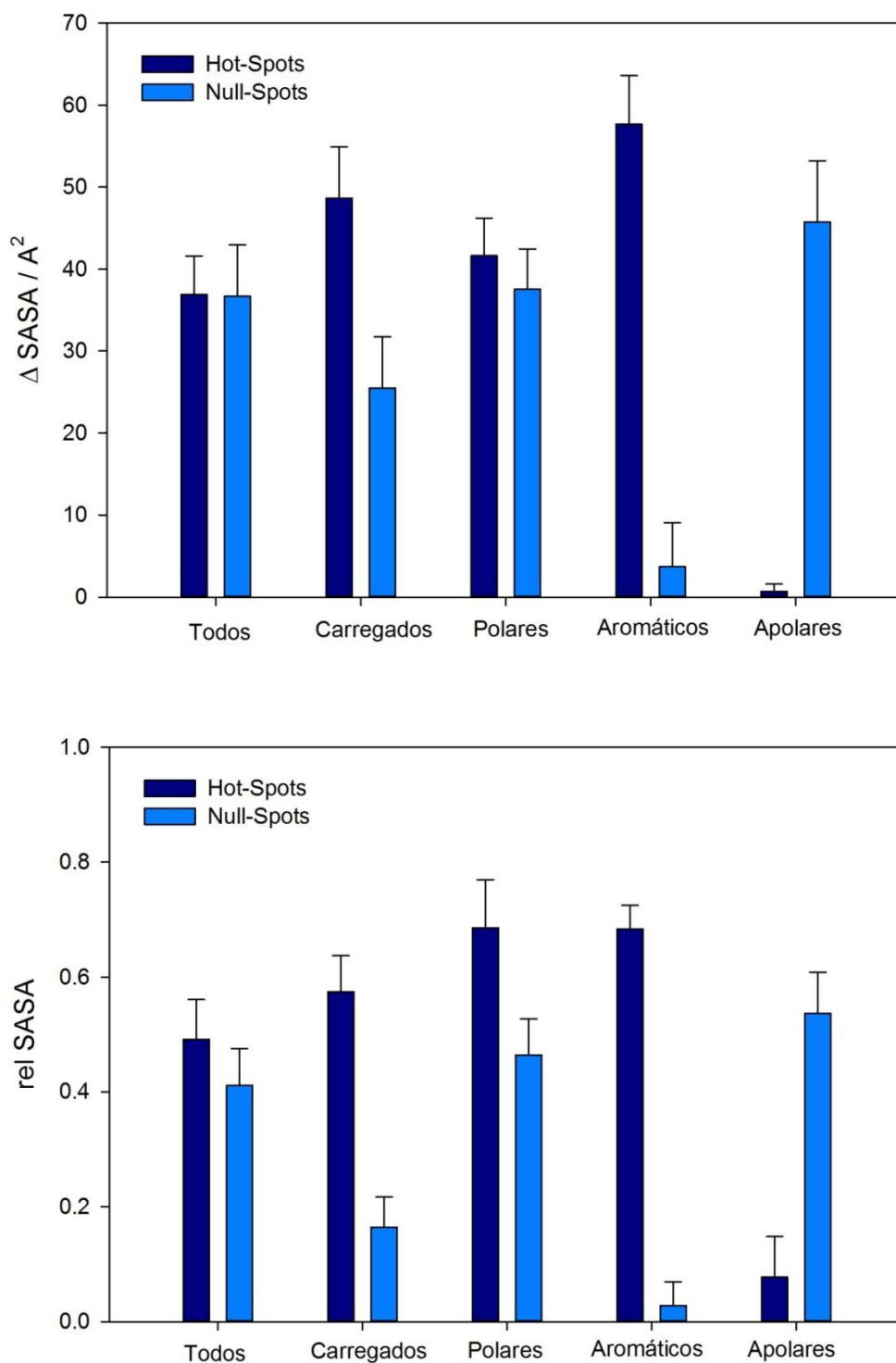


Figura III.7 - Representação esquemática dos valores de ΔSASA e relSASA obtidos para as 60 mutações analisadas.

2. MECANISMO DE INIBIÇÃO DO TENOFOVIR

Sendo a inibição da enzima HIV-1RT um passo crucial na luta contra a SIDA, o estudo das características estruturais e dinâmicas de um dos seus inibidores mais utilizados (TDF) pareceu-nos de grande importância. O principal foco do estudo centrou-se nos princípios que governam a excisão do TDF. Duas estruturas cristalográficas do TDF em complexo com a HIV-1 RT foram anteriormente descritas [83], sendo que na primeira estrutura o TDF encontra-se não covalentemente ligado à RT no N-site, seguido de terminação de cadeia de forma a impedir a sua incorporação no primer; na segunda estrutura o TDF encontra-se ligado covalentemente ao primer, no P-site. Nesta estrutura o TDF adota duas conformações diferentes, em que a adenina do TDF roda 180° entre as duas conformações (Figura III.8). Ainda permanece desconhecida a importância de fatores como a estrutura acíclica, o grupo fosfonato ou a presumível flexibilidade do TDF promovem a excisão ou a formação de um complexo final estável. Algumas hipóteses foram avançadas de forma a explicar estes fatores. Uma delas sugere que a flexibilidade da estrutura acíclica e as múltiplas conformações do TDF poderiam conferir resistência à excisão[83]. *Marchand et al.*[93] estudaram os fatores que governam a alteração no equilíbrio translacional, entre o complexo pré-translacional e o pós-translacional. O substrato natural (dNTP) tem um papel fundamental neste processo, tendo os autores verificado que a excisão do TDF é eficiente a baixas concentrações de dNTP, enquanto um aumento na concentração de dNTP (para valores que reproduzem a concentração na célula) levava à formação de um complexo terminal e consequente diminuição da excisão. Quando comparado o estado de translocação do TDF com o AZT, verificou-se que o AZT apresentava um rácio de (85:15 %) entre o estado pré-translacional e o estado pós-translacional, respetivamente, enquanto para o TDF era de (55:45 %). O rácio apresentado para o AZT não é de estranhar, uma vez que este apresenta dificuldades na formação de um complexo terminal devido a conflitos estéricos entre o seu grupo azido e o próximo dNTP. Quanto ao TDF, o facto de apresentar percentagens relativamente altas de excisão, apesar de menor acesso ao estado pré-translacional (N-site, necessário para excisão mediada por ATP), sugere a existência de algum fator a influenciar a eficiência de excisão.

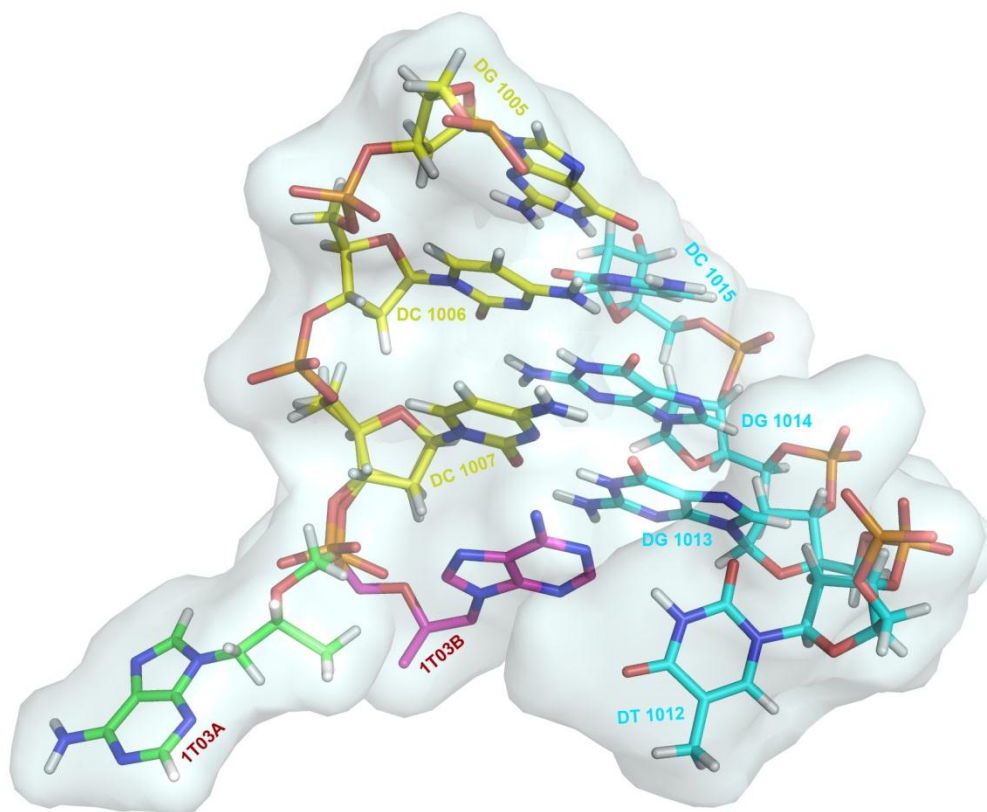
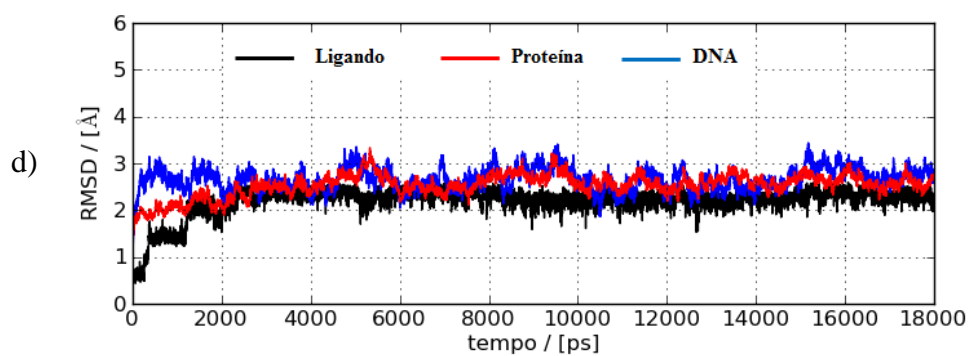
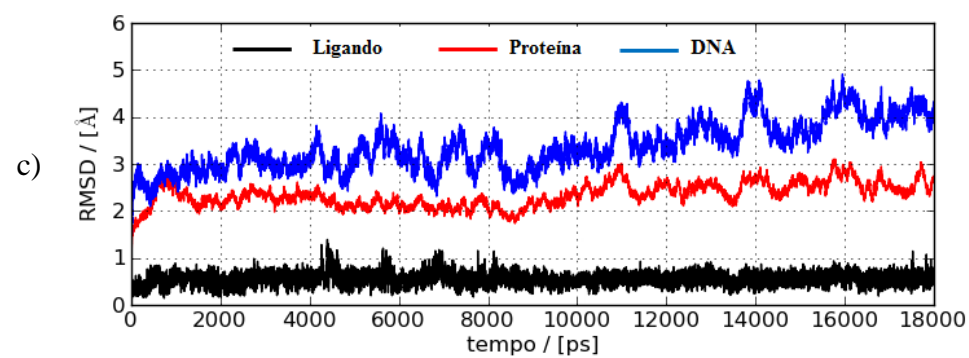
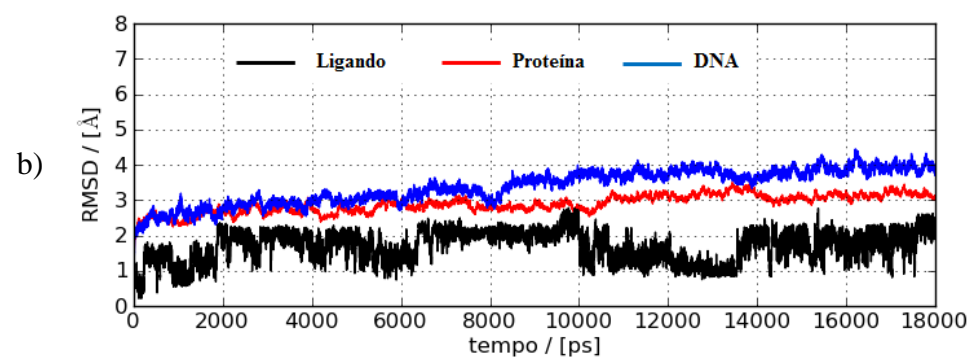
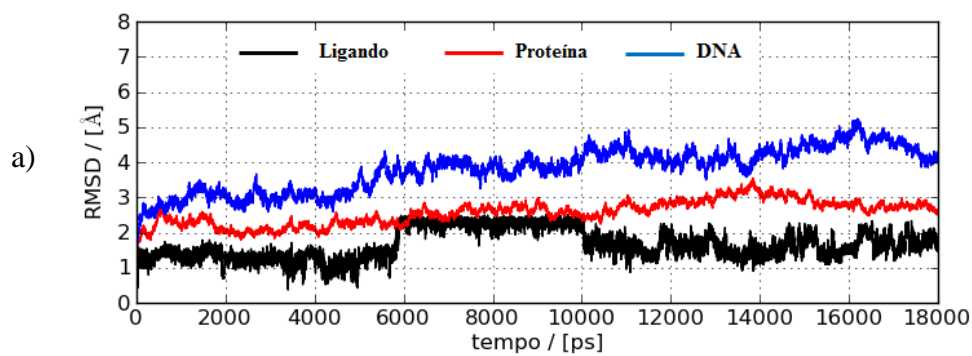


Figura III.8 - Representação do centro activo do complexo HIV-1 RT / DNA / TDF, no P-site (PDB ID: 1T03). TDF na conformação A (verde), TDF na conformação B (magenta), primer (amarelo), template (azul).

2.1. RMSD

A avaliação da estabilidade dos sistemas produzidos por DM é um passo essencial e que deve ser efetuado antes dos restantes, de forma a assegurar que os sistemas se encontram nas condições de equilíbrio desejadas. Para tal foi feita uma análise dos RMSD utilizando a estrutura inicial como referência. No cálculo dos RMSD considerou-se o backbone da proteína e do DNA (P, O3', O5', C3', C4' e C5'). De seguida (Figura III.9) apresentam-se os RMSD dos oito complexos em estudo. Com este estudo foi possível aferir da estabilidade dos sistemas ao longo das simulações de DM.



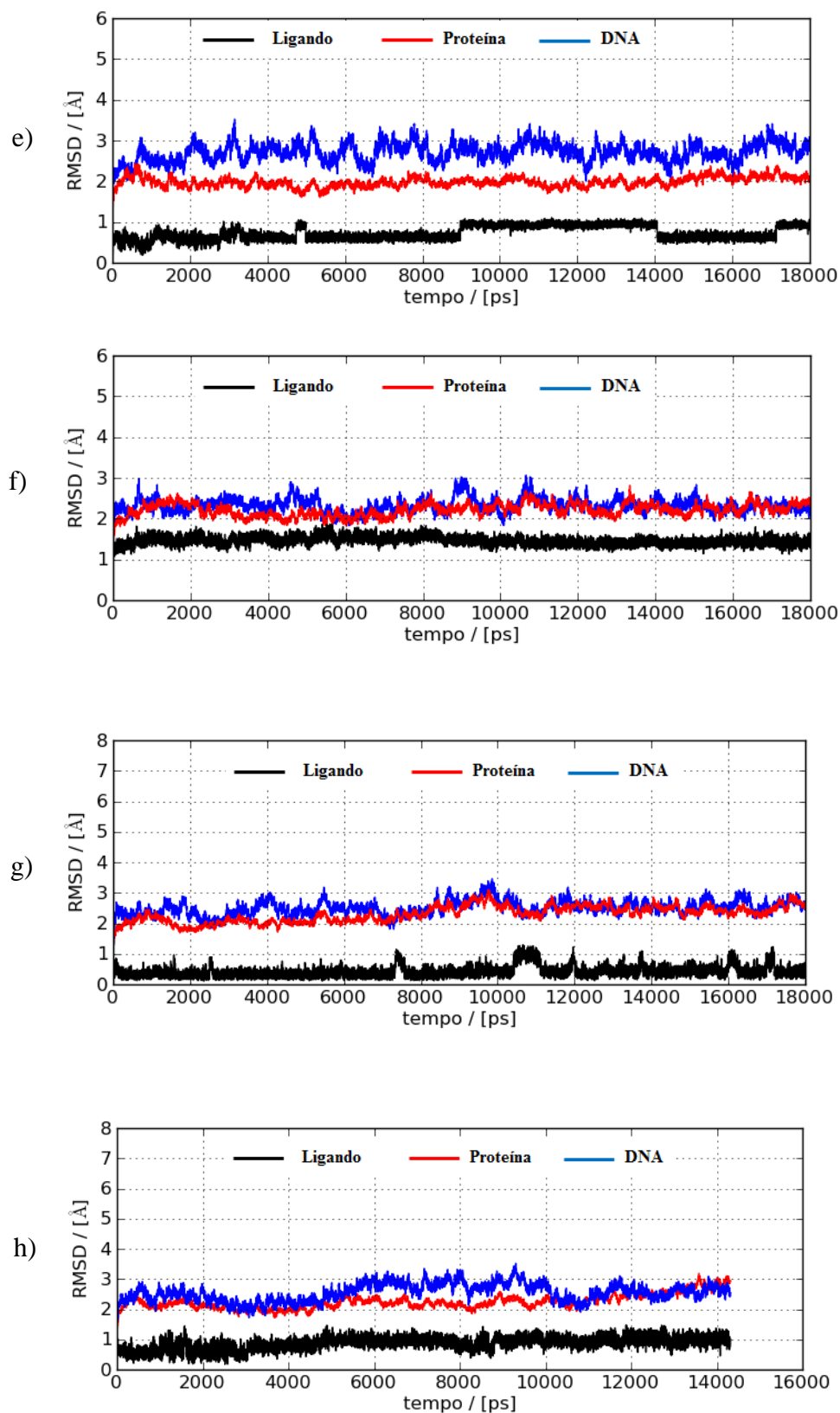


Figura III.9 - Representação da variação de RMSD da componente proteica, DNA e ligando para os sistemas: a) 1T03A; b) 1T03B; c) 1R0A; d) 1T05; e) 1RTD; f) 3JSM; g) 3JYT; h) 1NSY

2.2. Estudo da flexibilidade do TDF no P-site**2.2.1. Medição de ângulos diedros**

Como se referiu anteriormente, a estrutura cristalográfica 1T03 apresenta-nos o TDF em duas conformações distintas (1T03A e 1T03B). Juntamente com a percepção visual dos movimentos do TDF, utilizando a interface gráfica do VMD, procuraram-se utilizar outros métodos que nos permitissem ter uma caracterização exata do comportamento do TDF e das suas conformações ao longo da simulação. O nosso objetivo era o de verificar se o TDF permanecia constante ao longo da simulação, ou se por outro lado, adquiria facilmente outras conformações. Assim, para caracterizar os movimentos e interações realizados pelo TDF no centro ativo, começamos por medir uma variedade de ângulos diedros que, no nosso entender, nos poderiam dar informações sobre a forma como o TDF se movimentaria. Tendo em conta que na estrutura cristalográfica existe uma rotação da base adenina do TDF, em direcção ao interior do centro ativo, procuraram-se incluir diedros que caracterizassem esta rotação. Uma representação de alguns dos diedros medidos neste trabalho pode ser observada na Figura III.10. Os átomos que formam estes diedros não encontram correspondência direta naqueles que comumente são medidos para proteína e DNA, nomeadamente os diedros: Φ [C(O)-N-C(α)-C(O)], ψ [N-C(α)-C(O)-N] e ω [C(α)-C(O)-N-C(α)] (proteína), ε [C-C-O-P] e ζ [C-O-P-O] (DNA) [94]. Como tal, resolveu-se designar os ângulos diedros por Ω .

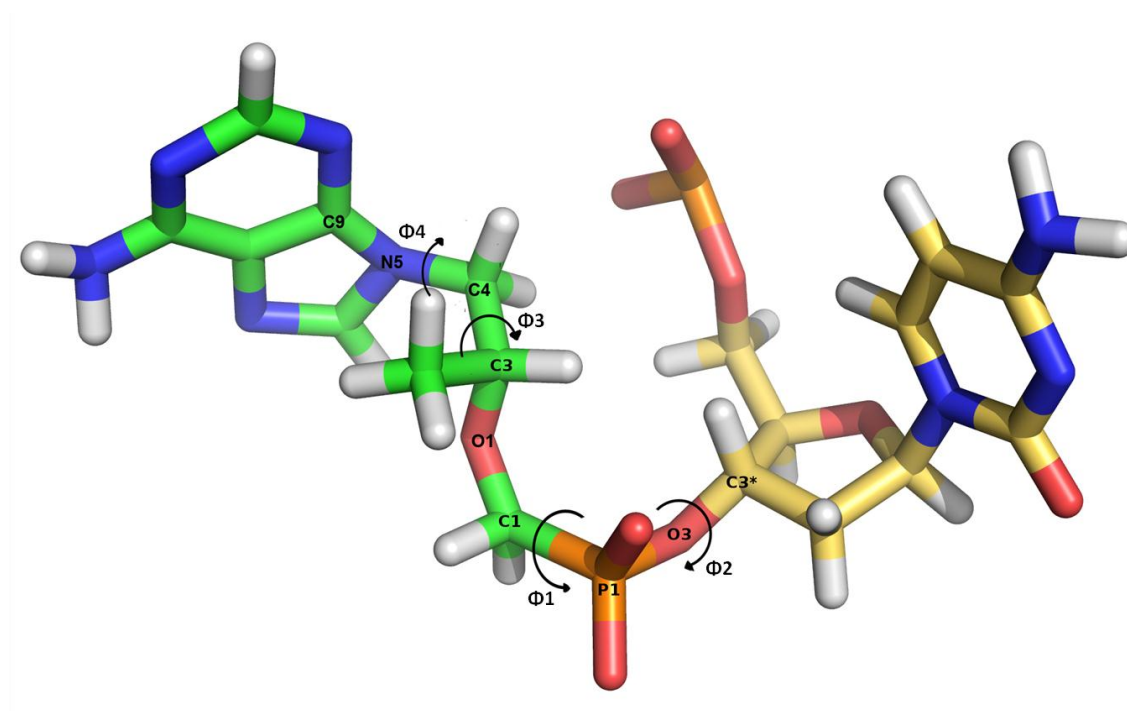


Figura III.10 - Representação de alguns dos ângulos diedros medidos neste trabalho: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3). TDF e dCMP estão coloridos a verde e amarelo, respectivamente.

A variação dos ângulos diedros ao longo da simulação de DM partindo da estrutura cristalográfica 1T03A, 1T03B e 1R0A pode ser conferido na Figura III.11, Figura III.12, Figura III.13, respetivamente.

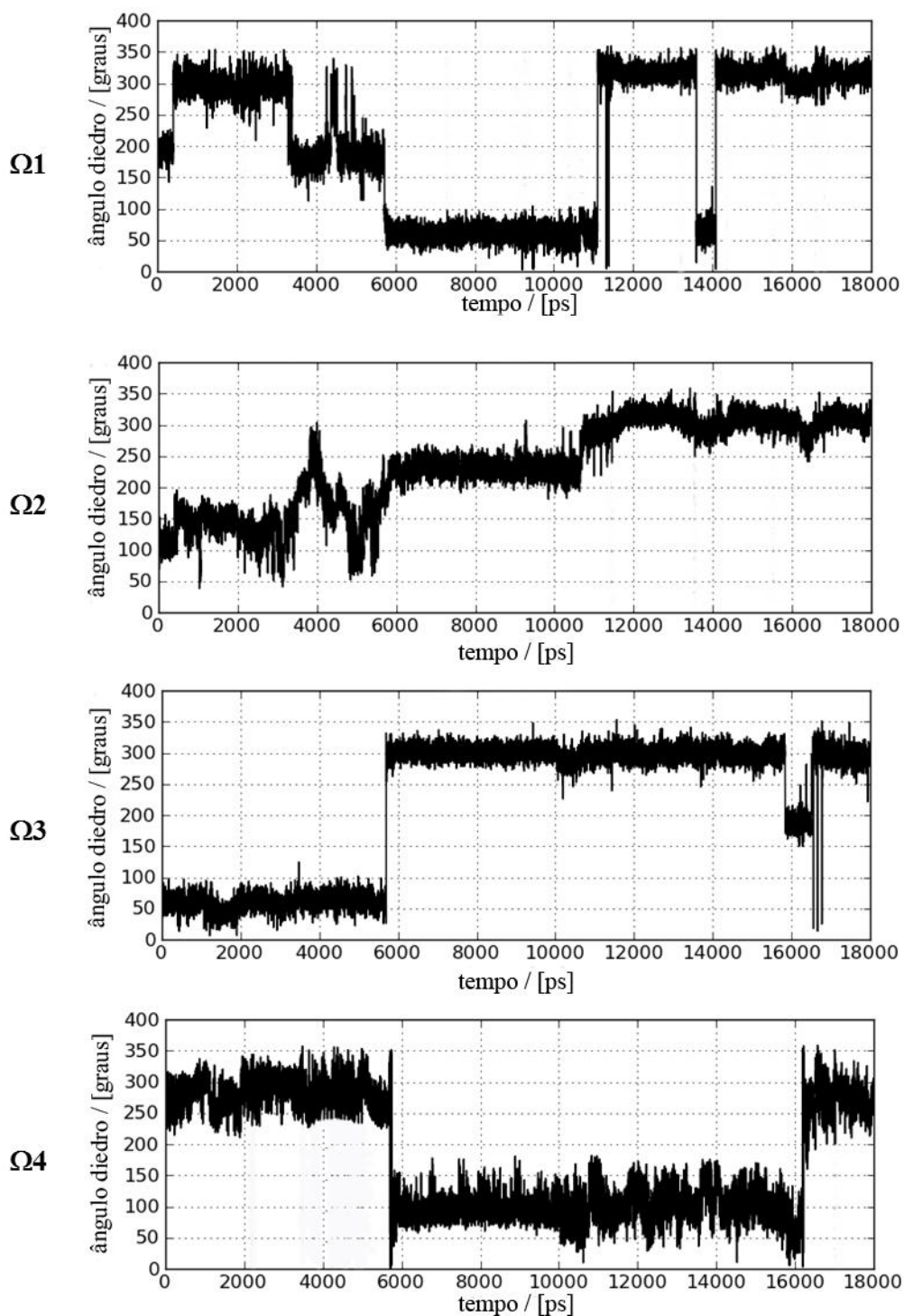


Figura III.11 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1T03A: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3).

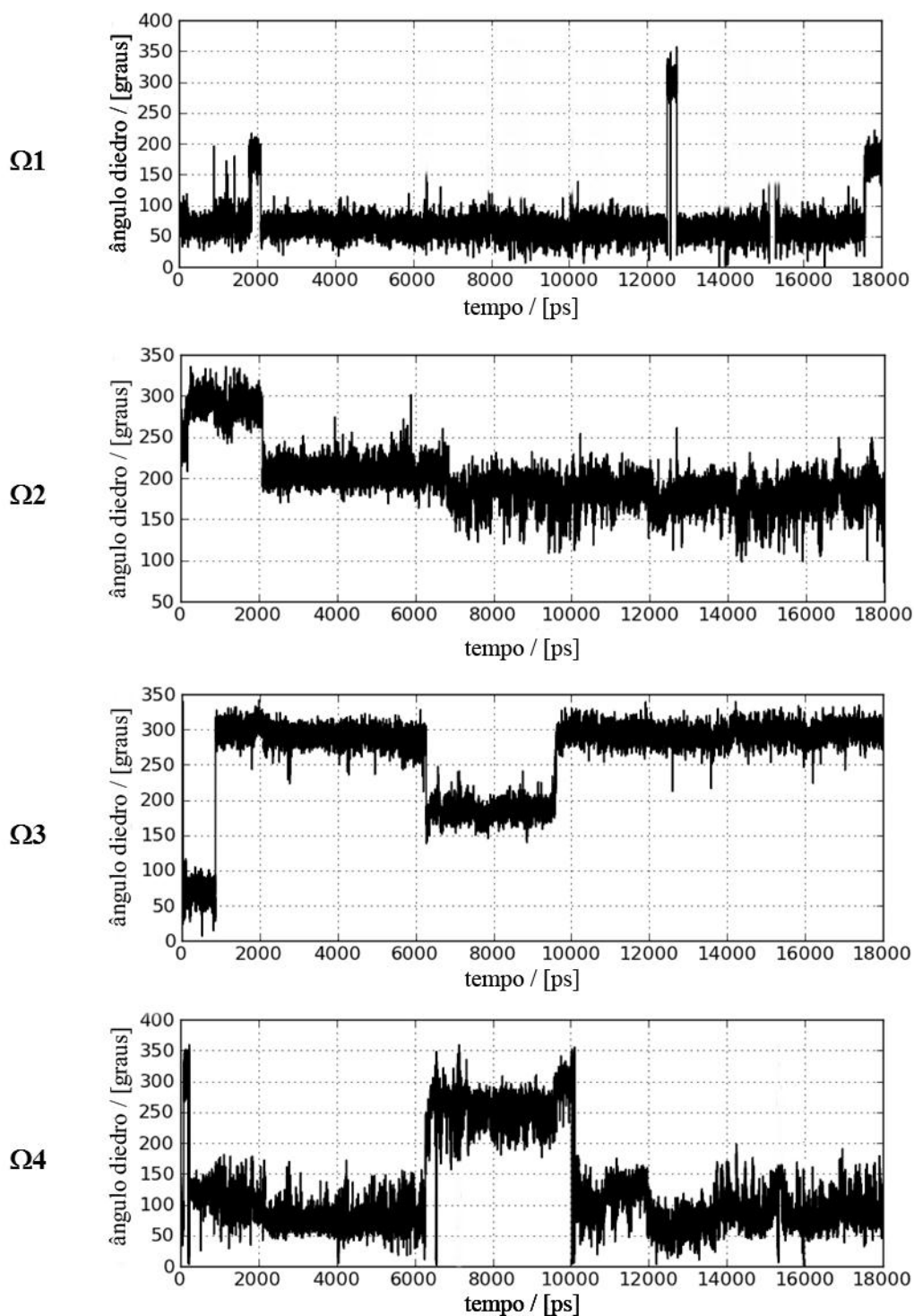


Figura III.12 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1T03B: Ω_1 (O1-C1-P1-O3'); Ω_2 (C1-P1-O3'-C3'); Ω_3 (N5-C4-C3-O1); Ω_4 (C9-N5-C4-C3).

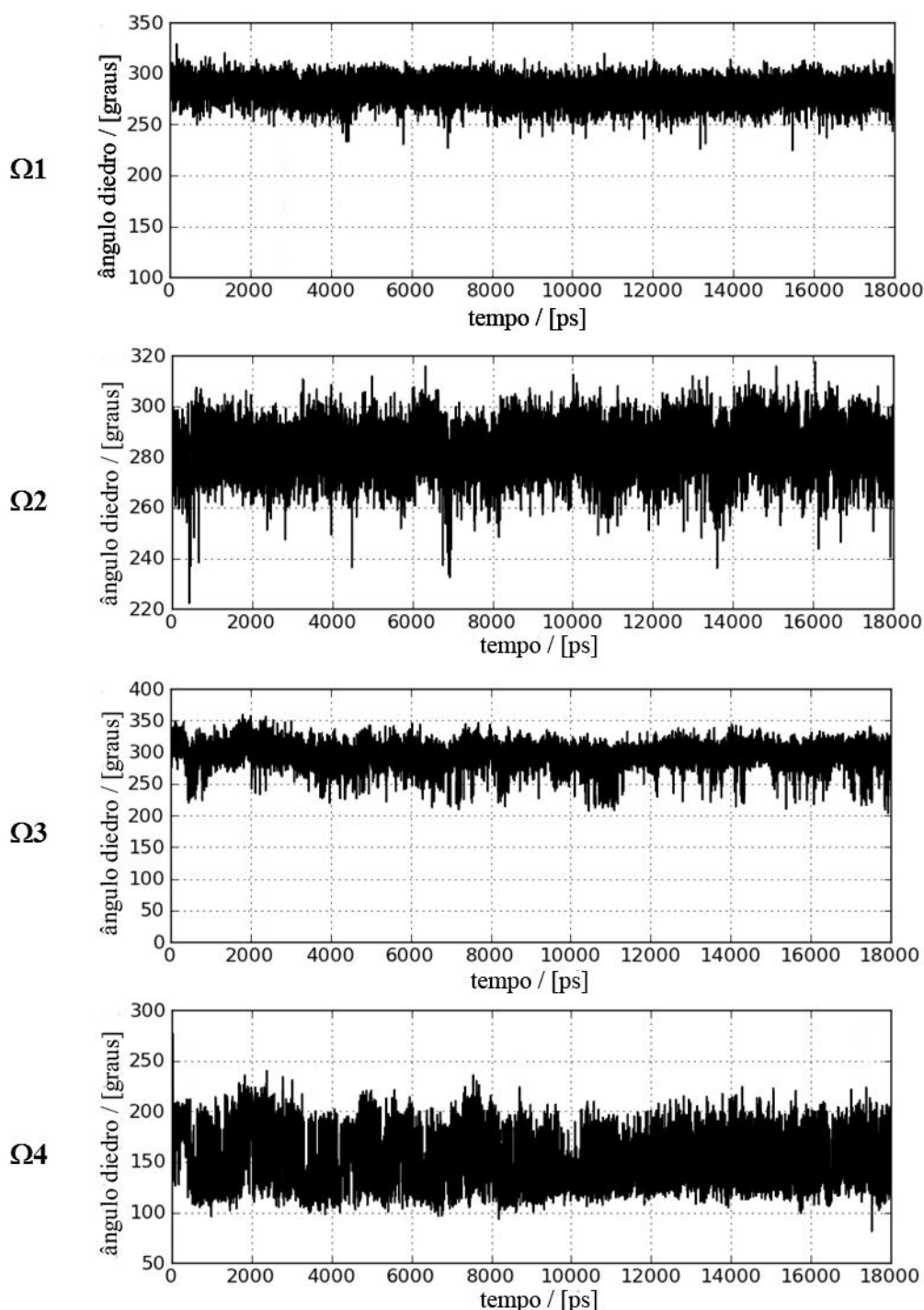


Figura III.13 - Representação da variação do ângulo diedro para os quatros casos considerados, partindo da estrutura cristalográfica 1R0A.

Fazendo uma análise superficial das figuras III.11, III.12 e III.13 é possível constatar de imediato a diferente variação nos ângulos diedros, em cada uma das estruturas cristalográficas analisadas. Considerando que os quatros ângulos diedros caracterizam, de forma aceitável, o movimento do ligando (TDF ou dATP) é possível dizer que existe uma diminuição da sua flexibilidade desde a estrutura cristalográfica 1T03A até à estrutura cristalográfica 1R0A, ou seja: $1T03A > 1T03B > 1R0A$. De facto, analisando a variação ocorrida para os ângulos

Capítulo III: Resultados e Discussão

diedros na Figura III.13, torna-se possível caracterizar o comportamento do ligando natural (dATP) como sendo praticamente constante. Desde o início até ao final da simulação não se encontra uma zona de variação brusca dos ângulos diedros, mantendo-se sempre dentro de um intervalo constante, o que nos pode indicar que desde o início até ao final da corrida, este ligando se manteve sempre na mesma conformação. O mesmo não pode ser dito no caso da simulação partindo da estrutura cristalográfica 1T03A (Figura III.11). Neste caso é possível identificar zonas de variação brusca do ângulo diedro, o que pode indicar uma alteração conformacional do ligando. Analisando de forma mais cuidada cada uma das representações na Figura III.11, e começando no diedro Ω_1 , é possível identificar uma zona no intervalo [0-6 ns] podendo ocorrer duas conformações (por volta dos 180° e 300°), outra zona no intervalo [6-11 ns] e outra no intervalo [11-18 ns]. O diedro Ω_2 tem o mesmo tipo de comportamento, ao passo que nos diedros Ω_3 e Ω_4 apenas são perceptíveis dois intervalos, [0-6 ns] e [6-18 ns]. O comportamento do TDF na estrutura cristalográfica 1T03B parece, de certa forma, ser mais constante do que no caso anterior, pois não se verificam oscilações tão acentuadas no valor do ângulo diedro. Ainda assim, conjugando a informação dos quatro ângulos consegue-se definir três intervalos diferentes, nomeadamente dos [0-6 ns], [6-10 ns] e [10-18 ns].

Apesar desta análise, realizada com recurso à medição de ângulos diedros, ser um bom ponto de partida para o trabalho que pretendíamos desenvolver, não é suficiente para caracterizar categórica e inequivocamente o movimento do TDF no centro ativo, ou as várias conformações que pode adoptar. Uma certa ambiguidade pode ocorrer e um ponto negativo desta análise baseada em ângulos diedros prende-se com o facto do movimento do TDF não ser devido, unicamente, à rotação de uma determinada ligação. Pelo contrário, deve-se à rotação de um, dois, três ou mais ângulos diedros que em conjunto caracterizam o seu movimento. Desta forma torna-se necessário encontrar outros métodos que nos permitam caracterizar de forma mais rigorosa o comportamento do TDF no centro ativo. A única exceção acaba por ser o ligando natural, que tendo um comportamento constante ao longo da simulação, consegue ser identificado através desta medição.

2.2.2. Cálculo de distâncias entre bases emparelhadas

A medição dos ângulos diedros não nos forneceu uma caracterização clara dos movimentos do TDF, pelas razões discutidas anteriormente. Tornou-se, pois, necessário encontrar métodos alternativos que, em complemento com a perceção visual, nos fornecessem informações claras sobre a flexibilidade do TDF no P-site. Uma característica particular do TDF chamou-nos à atenção, quando representamos apenas o TDF e o nucleótido (dCMP) a ele ligado no centro

ativo, pois constituía uma diferença clara entre as conformações da estrutura cristalográfica (Figura III.14).

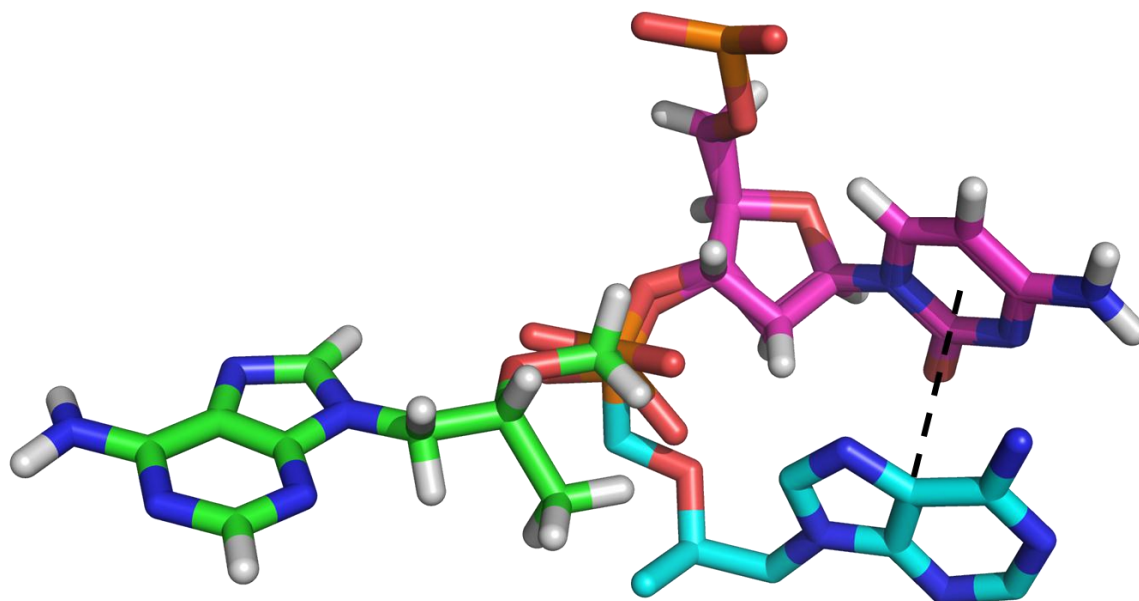


Figura III.14 - Representação da conformação inicial das estruturas cristalográficas 1T03A (verde) e 1T03B (azul). Na estrutura 1T03B o TDF encontra-se numa posição em que a sua base está praticamente paralela à base do resíduo dCMP (magenta).

O TDF, na estrutura cristalográfica 1T03B, apresenta uma orientação na qual a sua base adenina está praticamente paralela à base citosina do nucleótido emparelhado. Na conformação oriunda da estrutura cristalográfica 1T03A este paralelismo não acontece, pois o TDF encontra-se voltado para o exterior do centro ativo, tendo a sua base rodado cerca de 180° desde a conformação anterior. Esta característica pareceu-nos importante, pois a distância entre as bases seria um indicador da orientação do TDF, ou seja, se o TDF se encontrasse numa posição semelhante a 1T03B a distância entre as bases seria curta; por outro lado, se este se encontrasse numa conformação semelhante a 1T03A a distância entre as bases seria grande. Com base neste pressuposto resolvemos calcular o centro geométrico de ambas as bases (usando apenas os átomos pesados) e determinar a distância entre o centro geométrico de ambas as bases. Os resultados obtidos podem ser consultados na Figura III.15.

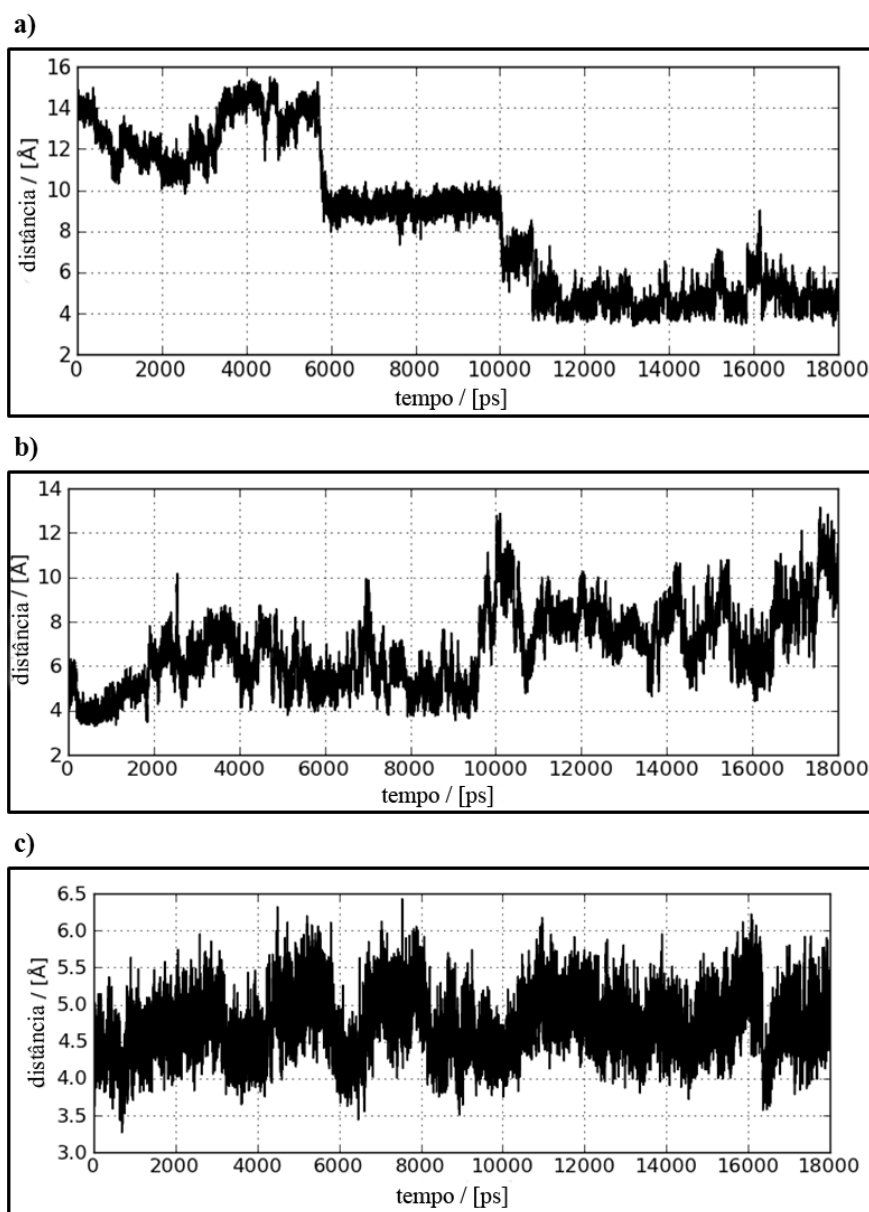


Figura III.15 - Representação da distância medida entre as bases adenina (TDF) e citosina (dCMP) para os complexos 1T03A (a), 1T03B (b) e 1R0A (c).

Com base na interpretação da distância entre as bases, os resultados para os ângulos diedros e a percepção visual dos acontecimentos ao longo da MD (utilizando o VMD) foi possível identificar as várias conformações do TDF na estrutura cristalográfica 1T03A e 1T03B.

Partindo da estrutura cristalográfica 1T03A, conseguem-se distinguir três conformações principais do TDF. O TDF não é um ligando estático, apresenta flexibilidade e capacidade para se movimentar, e por isso dentro destas três conformações podem ocorrer pequenas alterações conformacionais, que não são relevantes uma vez que ou acontecem num espaço de tempo muito curto (ps) ou são pequenos rearranjos que acabam por não ter importância na

globalidade da conformação. As três conformações podem ser definidas como: conformação anti-paralela (0-6 ns), conformação paralela (11-18 ns) e conformação intermédia (6-11 ns). Estas designações para as conformações foram baseadas no posicionamento da base do TDF relativamente à base do dCMP. Na Figura III.16 estão representadas as três principais conformações vistas na simulação de DM partindo da estrutura cristalográfica 1T03A. Na conformação que ocorre nos primeiros 6 ns da DM, o TDF encontra-se orientado para o exterior do centro ativo fazendo um ângulo de 180° com o dCMP e como tal numa conformação anti-paralela, que é igual à conformação inicial da estrutura cristalográfica 1T03A. O gráfico da distância entre as bases, e mesmo os ângulos diedros, mostram que existe até certo grau alguma variabilidade conformacional por parte do TDF, mas como referido anteriormente esta não é significativa, sendo possível considerar a conformação global. Nestes primeiros 6 ns existe uma distância média de $12,8 \pm 1,3$ Å entre a base adenina e a base citosina, confirmando a direção oposta em que estes dois resíduos se encontram. A conformação seguinte (intermédia, 6-11 ns) resulta do movimento em direção ao interior do centro ativo, ficando num estado intermédio entre a conformação anti-paralela e a conformação paralela. Ao contrário da conformação anterior, não se vislumbram variações conformacionais, sendo esta a única conformação ao longo dos 5 ns. A distância média entre as bases decresce para $9,2 \pm 0,4$ Å, denotando a aproximação que ocorreu entre ambas. Dos 11-18 ns o TDF adota uma conformação paralela, movendo-se para o interior do centro ativo e ficando lado-a-lado com o dCMP. Como seria de esperar, e como pode ser comprovado pela variação observada no gráfico a) da Figura III.15, existe uma diminuição da distância que separa as bases, tendo decrescido para um valor médio de $4,9 \pm 0,9$ Å.

Depois da caracterização das três conformações adotadas torna-se importante realçar o tipo de movimento demonstrado pelo TDF ao longo da simulação de DM. Começa numa conformação anti-paralela, passa por uma conformação intermédia e termina numa conformação paralela. Poderia em qualquer uma das fases haver um retorno à conformação de partida, mas aquilo que se verifica é que existe a tendência de se movimentar no sentido de adquirir uma conformação semelhante à inicial da estrutura cristalográfica 1T03B. Os efeitos deste tipo de movimento na flexibilidade e no mecanismo de excisão do TDF serão discutidos posteriormente.

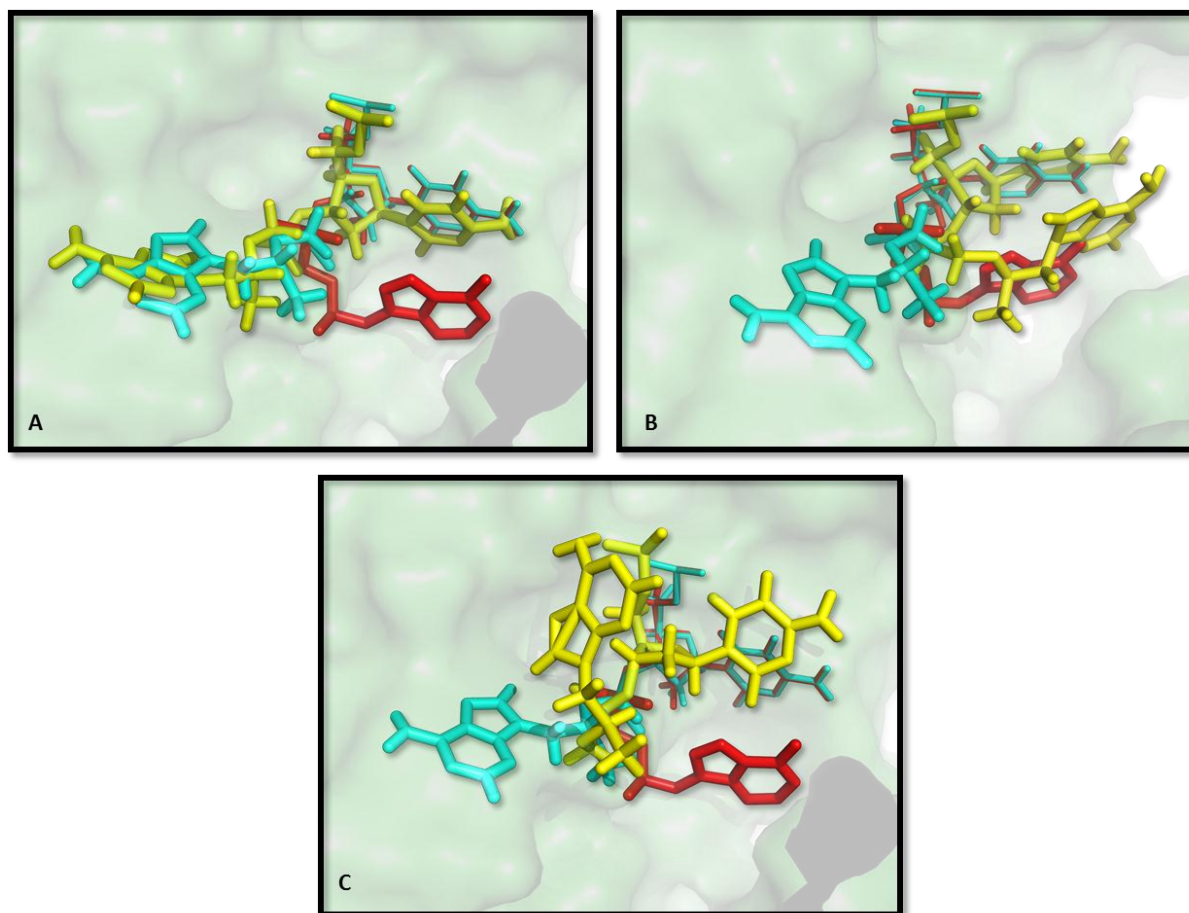


Figura III.16 - Representação das conformações adotadas pelo TDF na estrutura 1T03A no P-site (amarelo). As estruturas cristalográficas 1T03A e 1T03B estão representadas a azul e vermelho, respetivamente. A) Conformação anti-paralela; B) Conformação paralela; C) Conformação intermédia.

O mesmo estudo foi feito para a estrutura cristalográfica 1T03B, em que o TDF se encontra inicialmente numa conformação paralela. É importante referir que ao contrário do que acontece na estrutura 1T03A, em que todas as conformações apresentam grandes diferenças, na estrutura 1T03B a variação é muito menos acentuada. Para além de ser uma variação menos acentuada é uma variação que acontece numa zona restrita do centro ativo, uma vez que o TDF permanece praticamente sempre voltado para o interior e nunca adquire uma conformação semelhante à inicial 1T03A. Apesar da menor variabilidade conformacional exibida e da maior amplitude de movimentos que o TDF apresenta, definiram-se quatro zonas em que existe maior preponderância de uma determinada conformação. Assim, durante os primeiros 2 ns o TDF mantém-se praticamente inalterável em relação à conformação inicial, estando pois numa conformação paralela, com uma distância média entre as bases de $4,2 \pm 0,60$ Å (Figura III.15). De seguida (2-6 ns), o TDF parece adotar uma estrutura compacta, dobrado sobre si mesmo, fazendo lembrar a conformação intermédia vista em 1T03A mas mais

compacta. Na terceira zona (6-10 ns) existe o retorno a uma conformação paralela, sendo o aumento da distância entre as bases explicado por uma pequena movimentação do nucleótido dCMP. No último intervalo considerado (10 – 18 ns) verifica-se que o TDF apresenta uma flexibilidade acrescida, demonstrando uma grande amplitude de movimento à volta da conformação paralela. Este movimento é sobretudo vertical, no sentido em que não existe o retorno a uma conformação anti-paralela ou mesmo intermédia, existindo sim um aumento na distância entre as bases causada pelo deslocamento vertical do TDF.

O comportamento exibido pelo TDF é muito diferente daquele verificado para a estrutura 1T03A. Enquanto na estrutura 1T03A as conformações são bem definidas, não existindo praticamente variação, e se verifica um movimento de forma a adquirir uma estrutura paralela, no caso da estrutura cristalográfica 1T03B, que inicialmente já se encontra numa estrutura paralela, mantém este comportamento ao longo da simulação de DM. Como referido anteriormente, existe maior tendência do TDF oscilar à volta da conformação inicial mas não se verifica a passagem para uma conformação anti-paralela, mantendo-se ao longo da simulação direcionado para o interior do centro ativo.

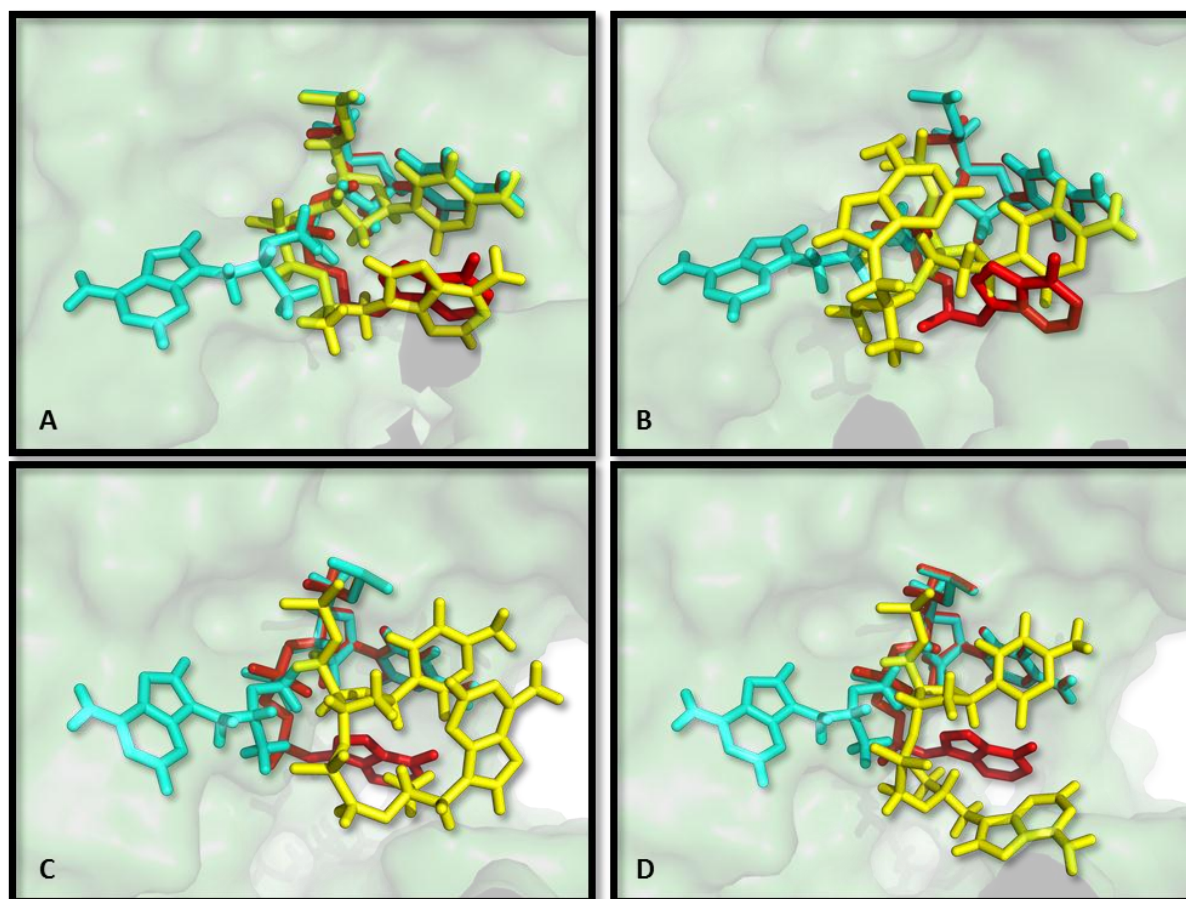


Figura III.17 - Representação das conformações adotadas pelo TDF partindo da estrutura 1T03B (amarelo). As estruturas cristalográficas 1T03A e 1T03B estão representadas a azul e vermelho, respectivamente. A) Conformação 0-2 ns; B) Conformação 2-6 ns; C) Conformação 6-10 ns; D) Conformação 10-18 ns.

Realizou-se o mesmo estudo para a estrutura cristalográfica 1R0A, ligando natural (dATP) no P-site. Este estudo serviria, principalmente, para comparar o comportamento do ligando natural com o do TDF ao longo da simulação de DM. Tal nos daria informação sobre o tipo de movimento do ligando natural no P-site, se permanecia voltado para o interior ou exterior do centro ativo, se adotava várias conformações ou se teria um comportamento constante. A análise aos ângulos diedros, como visto anteriormente, levava-nos a pensar que o ligando teria um comportamento constante, existindo maioritariamente numa conformação estável. Tal foi confirmado com a medição da distância entre as bases (neste caso duas adeninas), que como pode ser observado na Figura III.15-c se mantém constante, com um valor médio de $4,7 \pm 0,4$ Å. Quando comparada a conformação exibida pelo ligando natural com as estruturas cristalográficas 1T03A e 1T03B (Figura III.18), torna-se evidente a semelhança entre a conformação do ligando natural e a conformação 1T03B, ou seja, uma conformação paralela.

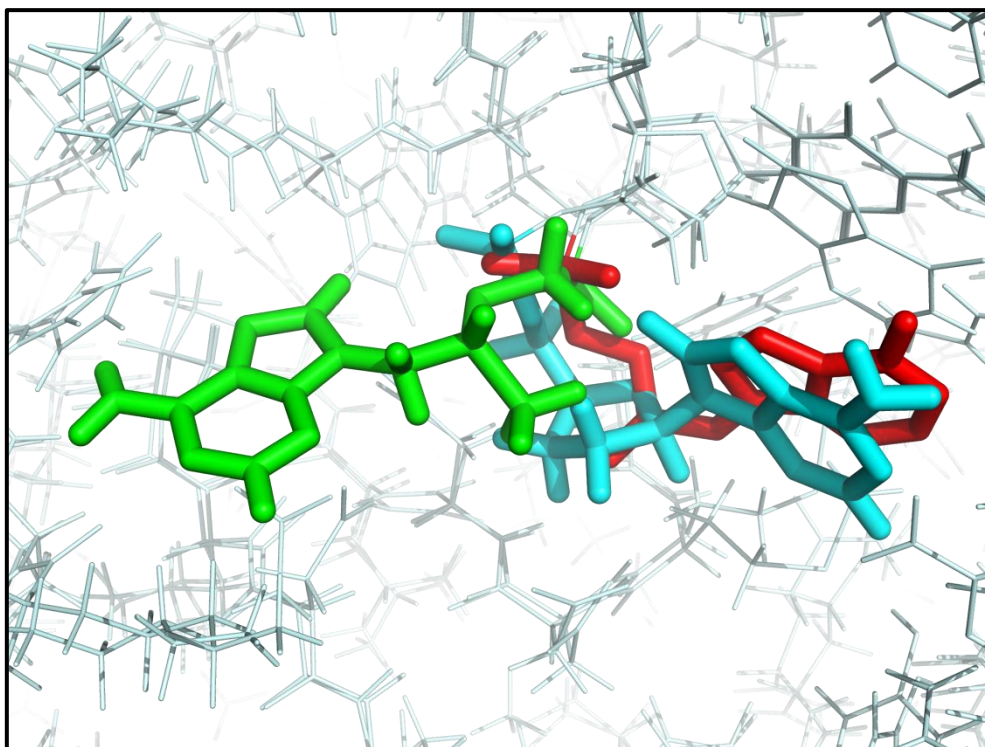


Figura III.18 - Representação das conformações iniciais das estruturas 1T03A (verde), 1T03B (azul) e 1R0A (vermelho).

Para complementar a informação obtida através da distância entre as bases do TDF e dCMP, e tendo em conta as limitações referidas anteriormente dos ângulos diedros, decidimos realizar novos cálculos e aplicação de métodos diferentes, para complementar a informação que já dispomos sobre as diferentes orientações do TDF. Este consistiria na medição do RMSD das bases (TDF + dCMP), utilizando como referência a estrutura de uma conformação identificada. Um exemplo da variação obtida para cada estrutura cristalográfica é apresentado de seguida. Identificam-se as três conformações da estrutura 1T03A, e as quatro na estrutura 1T03B, se bem que aqui com maior dificuldade devido à semelhança entre as conformações e à própria amplitude de movimento do TDF já referenciada.

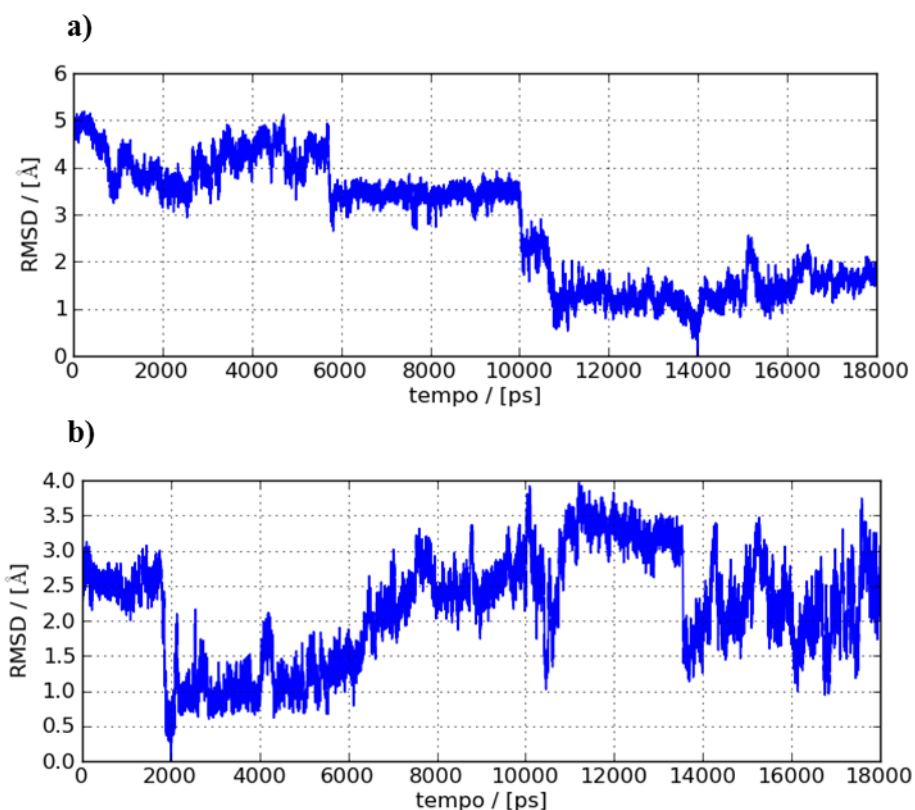


Figura III.19 – RMSD entre as bases (adenina e citosina) da estrutura 1T03A (a) e 1T03B (b).

2.2.3. Efeito da flexibilidade no mecanismo de excisão

Como foi discutido anteriormente, o TDF apresenta uma flexibilidade acentuada no P-site. Quando se compara o seu comportamento com o exibido pelo ligando natural denota-se um claro aumento da flexibilidade, uma vez que o ligando natural é bastante estável e adota sempre a mesma conformação ao longo da simulação. Os dados cristalográficos da estrutura tridimensional do TDF no P-site mostram-nos o TDF em duas conformações distintas. Neste trabalho, verificámos a existência de uma conversão, relativamente fácil, da conformação anti-paralela para a conformação paralela. Contudo, o contrário não se verifica, tendo a conformação paralela uma amplitude de movimentos em redor da conformação inicial, não se movimentando no sentido de adquirir uma conformação semelhante à anti-paralela. Esta situação levou-nos a colocar a hipótese da existência de alguma interação com o TDF, que o manteria naquela posição ao longo da DM. Tal interação pode envolver o emparelhamento entre as bases do TDF e do resíduo dCMP, bem como o contacto próximo com o DNA, que

apenas se faz sentir nesta conformação, “fixando” o TDF numa conformação direcionada para o interior do centro ativo.

Comparando a estrutura do TDF com a estrutura de outro inibidor da HIV-1 RT no P-site, o AZT (Figura III.20), verificamos que este apresenta uma conformação semelhante à estrutura 1T03B, ou seja, com a adoção de conformação paralela. Este comportamento é constante ao longo do tempo de simulação analisado, não existindo alterações significativas da conformação do ligando. A distância média entre as bases é de $4,90 \pm 0,63 \text{ \AA}$ que, tal como no Tenofovir, é um indicativo do emparelhamento existente.

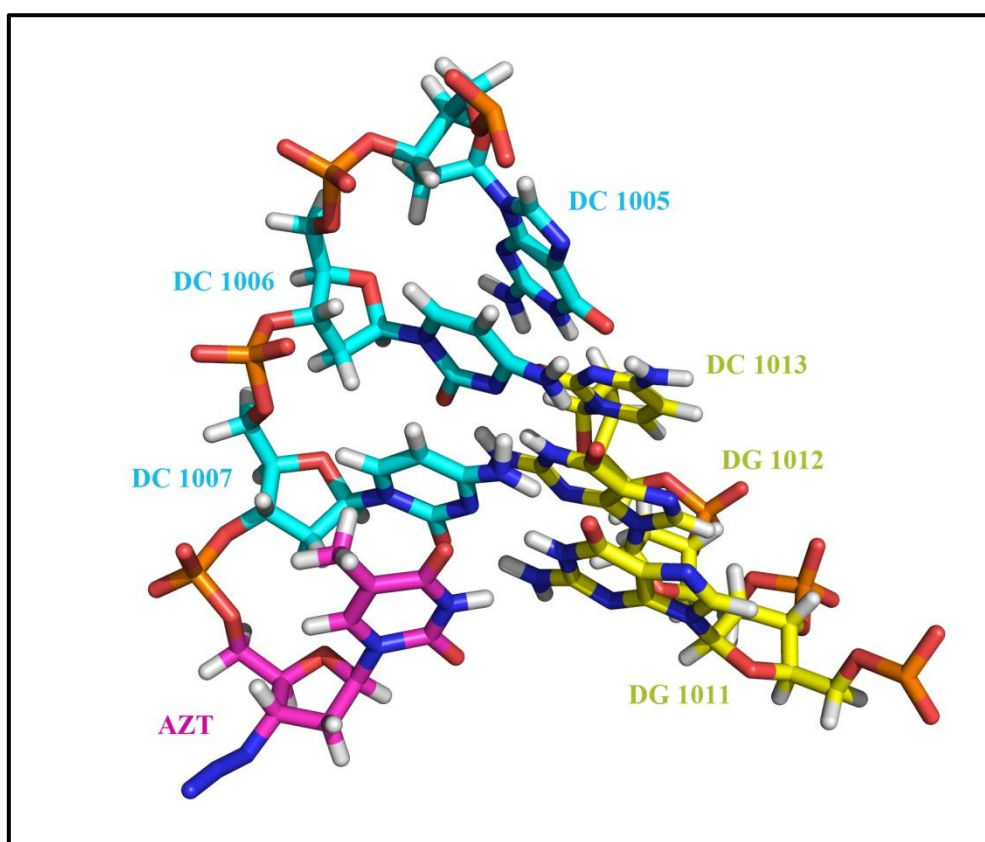


Figura III.20 - Representação do AZT (magenta) no centro ativo, primer (azul) e template (amarelo).

A excisão mediada por ATP só pode ocorrer no complexo pré-translacional, quando a ponta 3' do primer se encontra localizada no N-site. Por outro lado, a ligação do próximo nucleótido complementar só pode ocorrer se o N-site se encontrar livre no complexo pós-translacional, levando à formação de um complexo final estável que impede a excisão. Encontra-se, atualmente, bem documentada a dificuldade do AZT em formar um complexo final estável, resultando numa maior aptidão para o estado pré-translacional o que leva à excisão. A

Capítulo III: Resultados e Discussão

difficuldade do AZT em formar um complexo final estável resulta de conflitos entre o grupo azido e o substrato dNTP. Quanto ao TDF, a presença de um linker acíclico no lugar do grupo azido volumoso confere-lhe maior possibilidade de formar um complexo final estável. A flexibilidade aqui demonstrada poderá conferir ao TDF maior capacidade para se adaptar à incorporação do nucleótido, levando a emparelhamento estáveis entre a base do TDF, a base do dCMP e a base do nucleótido ligante. Por outro lado, e na ausência de formação de um complexo final, o facto de o TDF poder adquirir várias conformações (especialmente a conformação anti-paralela) poderá dificultar a retrotranslocação ao N-site e a um estado pré-translacional, diminuindo assim a percentagem de excisão.

De forma a confirmar as hipóteses aqui levantadas seria interessante estudar a possibilidade de modelar um sistema ligando um nucleótido ao TDF e observar o tipo de interações a ocorrer. Contudo, no âmbito deste trabalho e tendo em conta o tempo disponível para a sua realização, tal não foi possível.

2.3. Comparação entre a mutação K65R e a estrutura Wild-Type

Tal como se referiu anteriormente, o processo de síntese de DNA pela HIV-1 RT é propenso à aparição de mutações que causam resistência aos fármacos utilizados no combate à infeção. A mutação mais vezes associada a resistência ao TDF é a K65R, em que existe a mutação de uma lisina na posição 65 para uma arginina. Esta substituição adiciona um grupo maior à região do centro ativo, com capacidade para formar um maior número de interações (pontes de hidrogénio, etc.) com os resíduos circundantes. Segundo a bibliografia disponível, o mecanismo pelo qual a mutação causa resistência ao TDF deve-se ao emparelhamento das cadeias laterais dos resíduos Arg65 e Arg72, que impossibilitam a correta orientação do TDF para a ligação. Afeta, pois, o mecanismo de incorporação. Existe também a convicção que este emparelhamento, após a ligação do TDF, fixa o TDF numa conformação que diminui a excisão. No estudo da mutação K65R foram utilizados quatro dos sistemas anteriormente descritos: 3JSM e 3JYT, sistemas contendo a mutação em que o ligando é o TDF e o ligando natural dATP, respetivamente; 1T05 e 1RTD, estruturas *wild-type* (WT) com o TDF e o ligando natural dTTP, respetivamente. A estabilidade destes sistemas foi discutida anteriormente, tendo sido comprovada a sua estabilidade ao longo da DM (Figura III.9). O emparelhamento das cadeias laterais dos aminoácidos Arg65 e Arg72 foi confirmado para a estrutura cristalográfica mutante (3JSM), e como seria de esperar, o mesmo não foi verificado para a estrutura WT (1T05). Este emparelhamento só é possível devido à substituição da lisina pela argina, que possibilita a entrada de um grupo guanidina no centro ativo. Esta

substituição obriga o TDF a adotar uma conformação relativamente mais alongada (Figura III.21). Contudo, na simulação de DM efetuada para o complexo 3JSM, observou-se que o emparelhamento das cadeias laterais não é tão evidente como o demonstrado na estrutura cristalográfica. Existe uma aproximação entre a carga positiva da cadeia lateral da Arg72 e a carga negativa dos oxigénios do TDF, o que resulta num aumento da distância entre as cadeias laterais das duas argininas. Na estrutura cristalográfica esta distância era de 4,31 Å enquanto na nossa simulação obteve-se um valor médio de $7,9 \pm 1,2$ Å (contabilizando todos os átomos do grupo guanidina e medindo a distância entre o centro geométrico). Apesar do emparelhamento entre as duas cadeias laterais não ser tão efetivo, a forte interação entre a Arg72 e a base do TDF e entre a Arg65 e o grupo fosfato mantém-se. Ao longo dos 18 ns de simulação tanto a Arg72 como a Arg65 permanecem a uma distância inferior a 5 Å do TDF (100 % tempo ocupação).

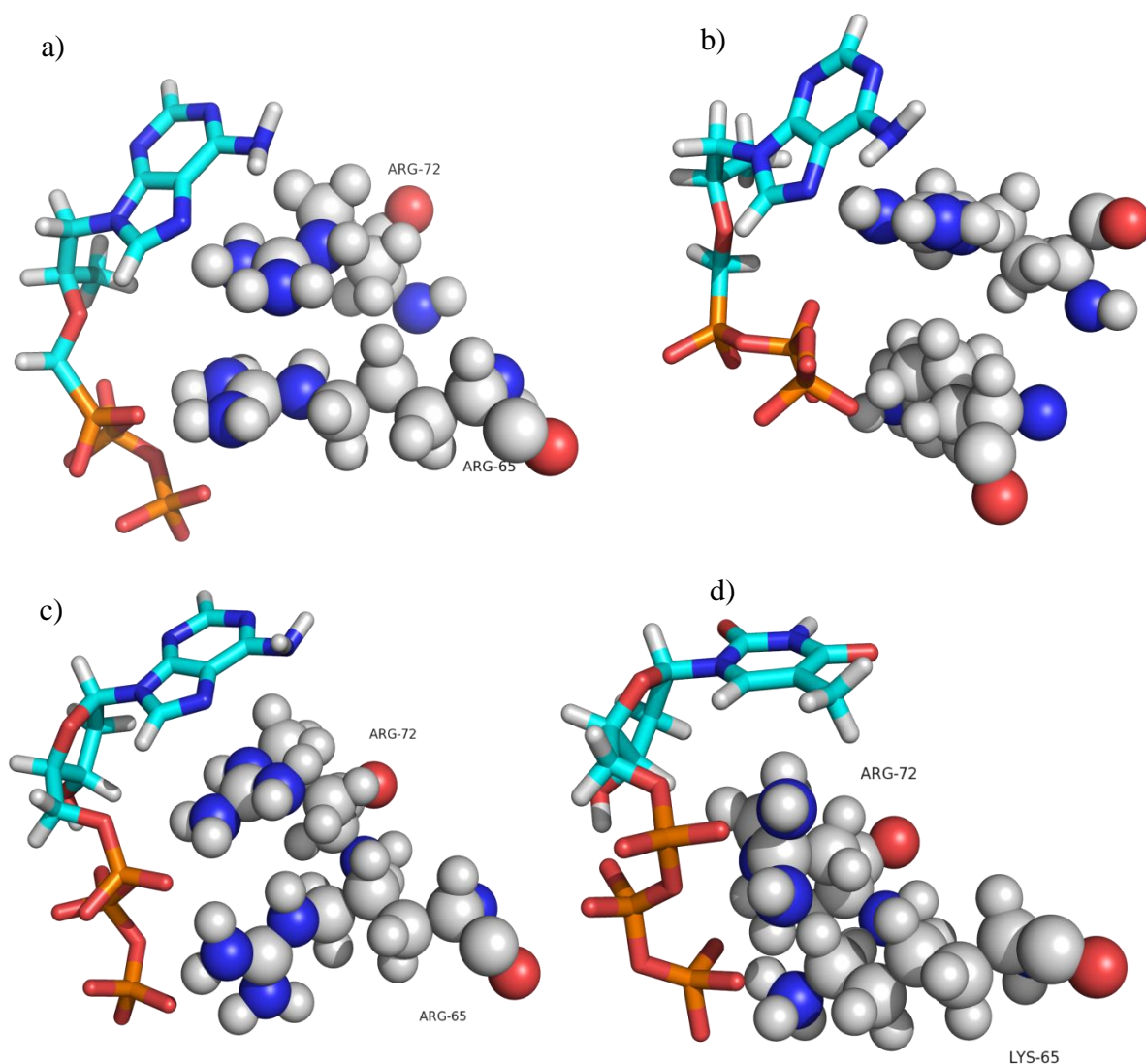


Figura III.21 - Representação da interação entre os resíduos Arg65, Arg72 e Ligando na estrutura 3JSM (a), 1T05 (b), 3JYT (c) e 1RTD (d).

Capítulo III: Resultados e Discussão

Tabela III.17 - Resultados da distância média e tempo de ocupação entre o Ligando e a Lys/Arg65 e Arg72.

PDB ID: Ligando	Mutação		Dist L:X65 / Å	Tempo Ocupação	Dist L:R72 / Å	Tempo Ocupação
1T05	TDF	Não	$6,97 \pm 2,31$	54,70%	$5,44 \pm 0,90$	99.30%
1RTD	dTTP	Não	$3,00 \pm 0,76$	100%	$2,55 \pm 0,80$	100%
3JSM	TDF	Sim	$3,87 \pm 0,97$	100%	$3,14 \pm 0,41$	100%
3JYT	dATP	Sim	$1,96 \pm 0,32$	100%	$3,27 \pm 1,12$	100%

O comportamento observado para a simulação de DM da estrutura 1T05 é completamente diferente, em grande parte devido à não mutação da lisina para arginina. A cadeia lateral da lisina é menos volumosa e tem menor capacidade de formar ligações de hidrogénio com resíduos na vizinhança. Aquilo que se observou foi a tendência da Lys65 se afastar da esfera de coordenação do TDF, e simultaneamente o movimento dos fosfatos em direção ao grupo guanidina da Arg72. Esta atração obriga o TDF a adotar uma conformação ligeiramente diferente, mais fechado e em torno da Arg72 (Figura III.22). Ao longo da simulação apenas em 54,70 % do tempo a Lys65 se encontra a menos de 5 Å do TDF, em contraste com os 100 % na estrutura 3JSM. A estrutura 3JYT apresenta um comportamento semelhante à da sua homóloga, sendo que neste caso a distância média entre os grupos guanidina é de $4,50 \pm 0,51$ Å. A simulação de DM da estrutura *wild-type* com o ligando natural (1RTD), interessantemente, apresenta um comportamento diferente do verificado para a estrutura *wild-type* com o TDF. A Lys65 apresenta-se mais próxima do ligando, fixando o grupo fosfato que assim já não se orienta em direção à Arg72. Foi com alguma surpresa que se constatou que ao longo dos 18 ns de simulação a Arg65 se mantém na esfera de coordenação do ligando (100 % tempo ocupação), contrastando com o comportamento observado na estrutura com o TDF. Um estudo mais aprofundado sobre os aminoácidos e moléculas de água presentes na esfera de coordenação do ligando, Arg72 e Lys/Arg65 seria interessante e poderia ajudar a esclarecer o comportamento da Lys65 na estrutura 1T05.

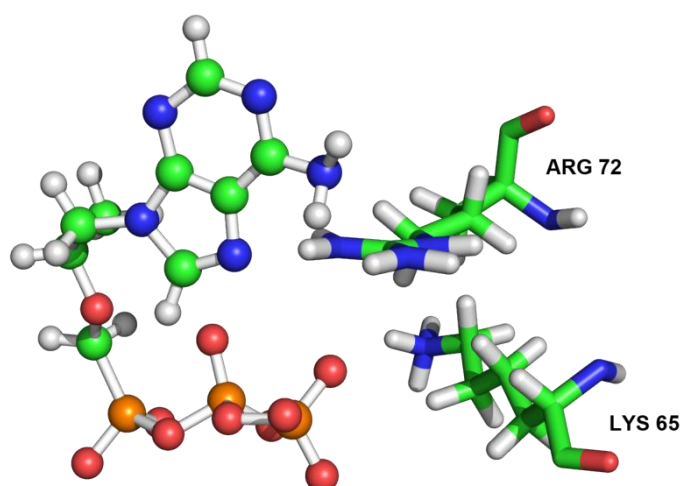


Figura III.22 - Representação da orientação conformacional do TDF, Arg72 e Arg 65 na simulação de DM da estrutura 1T05. Imagem retirada aos 4 ns.

Capítulo IV. Conclusão e Trabalho Futuro

CAPÍTULO IV: CONCLUSÃO E TRABALHO FUTURO

Neste trabalho pretendeu-se avaliar os efeitos da mutagénese na determinação de HS e NS num conjunto de complexos proteína-DNA por aplicação da metodologia computacional *Alanine Scanning Mutagenesis*, bem como estudar um dos mais importantes complexos formados entre proteínas e DNA: HIV-1 RT.

A determinação de HS e NS e o cálculo da energia livre de ligação para interfaces proteína-DNA é uma temática ainda pouco desenvolvida. Como tal aplicou-se a metodologia computacional ASM a cinco sistemas proteína-DNA num total de sessenta mutações. Diferentes tipos de parâmetros foram analisados e avaliados: i) tempo de simulação; ii) representação do solvente; iii) número de trajetórias; iv) força iónica e v) constante dielétrica interna. Os resultados obtidos demonstraram que o tempo de simulação (utilizou-se trajetórias de 2 e 6 ns em solvente explícito) provocava diferenças mínimas no cálculo da energia livre, sendo que a utilização de um tempo de simulação maior levava a tempos de computação maiores. O tipo de representação do solvente influencia claramente os resultados obtidos, existindo preferência por solvente explícito, o que é compreensível quando se tem em conta que o meio é constituído maioritariamente por resíduos carregados e governado por interações eletrostáticas. Os resíduos carregados constituíram um dos principais problemas na aplicação do método, pois originavam valores excessivamente elevados de energia livre. A solução encontrada passou pela utilização da equação não-linear de Poisson-Boltzmann, integrada no programa DelPhi, que como tem em consideração a força iónica é mais adequada para lidar com este tipo de resíduos e suas interações. O erro médio para o conjunto das 60 mutações e das 10 constantes dielétricas passa de 46,12 para 4,10 kcal mol⁻¹ com a alteração da equação linear para a equação não-linear, respectivamente. Em alguns casos, maioritariamente para os resíduos carregados, verificou-se que a melhor concordância com os valores experimentais era obtida apenas contabilizando a contribuição da energia eletrostática e de solvatação polar.

Para cada grupo de resíduos, nomeadamente carregados, polares e apolares, com base nos resultados obtidos definiram-se aquelas que seriam as condições mais favoráveis para a sua determinação, sendo que para os carregados seria a utilização de uma constante dielétrica 5, pelo método $\Delta\Delta E_{\text{ele}} + \Delta\Delta G_{\text{solv polar}}$; para resíduos polares e apolares a medição de $\Delta\Delta G_{\text{MM-PBSA}}$ com constantes dielétricas de 6 e 2, respetivamente. Com esta medição obteve-se um erro médio de 2,86 kcal mol⁻¹ e resultados estatísticos de 43%, 60%, 50%, 60% e 60% para o VPP, Sensibilidade, F1, Precisão e Especificidade, respetivamente. Globalmente podemos atestar

Capítulo IV: Conclusão e Trabalho Futuro

da capacidade da metodologia desenvolvida na determinação de HS e NS com boas taxas de sucesso, ainda mais se tivermos em conta que esta metodologia computacional nunca foi aplicada a sistemas proteína-DNA.

De futuro, seria interessante aplicar esta metodologia a um número maior de complexos, de preferência com um maior número de HS experimentais, bem como estudar a influência de diferentes concentrações iónicas na medição.

Relativamente ao estudo envolvendo a HIV-1 RT, confirmou-se a flexibilidade demonstrada pelo inibidor TDF no P-site que resulta na obtenção de múltiplas conformações. Na simulação de DM efetuada partindo da estrutura cristalográfica 1T03A verificou-se que o TDF adoptava três conformações distintas, com tendência para se movimentar em direção ao interior do centro ativo, para uma conformação semelhante ao ligando natural e a um outro inibidor da HIV-1 RT, o AZT. O facto do TDF poder existir numa conformação diferente da “habitual” e ao ter a capacidade de se movimentar e de alterar de conformação pode ajudar a explicar o facto deste inibidor ser menos excisado que os restantes inibidores. Quanto à mutação K65R verificou-se que, ao contrário do pressuposto inicial, não existe emparelhamento das cadeias laterais dos aminoácidos Arg72 e Arg65. Contudo a interação com o TDF é mantida, reduzindo-lhe a mobilidade conformacional necessária para se integrar na HIV-1 RT. Por outro lado, o TDF na estrutura WT, sem mutação, apresenta maior mobilidade conformacional uma vez que apenas sofre a interação direta da Arg72.

Capítulo V. Referências

1. Davies DR, Cohen GH: **Interactions of protein antigens with antibodies.** *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(1):7-12.
2. Moreira IS, Fernandes PA, Ramos MJ: **Hot spots-a review of the protein-protein interface determinant amino-acid residues.** *Proteins-Structure Function and Bioinformatics* 2007, **68**(4):803-812.
3. Lo Conte L, Chothia C, Janin J: **The atomic structure of protein-protein recognition sites.** *Journal of Molecular Biology* 1999, **285**(5):2177-2198.
4. Thorn KS, Bogan AA: **Aseddb: A database of alanine mutations and their effects on the free energy of binding in protein interactions.** *Bioinformatics* 2001, **17**(3):284-285.
5. Jones S, Thornton JM: **Principles of protein-protein interactions.** *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(1):13-20.
6. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces.** *Journal of Molecular Biology* 1998, **280**(1):1-9.
7. Moreira IS, Martins JM, Ramos RM, Fernandes PA, Ramos MJ: **Understanding the importance of the aromatic amino-acid residues as hot-spots.** *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* (0).
8. Moreira IS, Fernandes PA, Ramos MJ: **Detailed microscopic study of the full zipa : Ftsz interface.** *Proteins-Structure Function and Bioinformatics* 2006, **63**(4):811-821.
9. Moreira IS, Fernandes PA, Ramos MJ: **Unraveling the importance of protein-protein interaction: Application of a computational alanine-scanning mutagenesis to the study of the igg1 streptococcal protein g (c2 fragment) complex.** *Journal of Physical Chemistry B* 2006, **110**(22):10962-10969.
10. Sauereriksson AE, Kleywegt GJ, Uhl M, Jones TA: **Crystal-structure of the c2 fragment of streptococcal protein-g in complex with the fc domain of human-igg.** *Structure* 1995, **3**(3):265-278.
11. Moreira IS, Fernandes PA, Ramos MJ: **Computational alanine scanning mutagenesis - an improved methodological approach.** *Journal of Computational Chemistry* 2007, **28**(3):644-654.
12. Luscombe NM, Austin SE, Berman HM, Thornton JM: **An overview of the structures of protein-DNA complexes.** *Genome biology* 2000, **1**(1):REVIEWS001.
13. Norambuena T, Melo F: **The protein-DNA interface database.** *Bmc Bioinformatics* 2010, **11**.
14. Evertts AG, Zee BM, Garcia BA: **Modern approaches for investigating epigenetic signaling pathways.** *J Appl Physiol* 2010, **109**(3):927-933.
15. Ahmad S, Keskin O, Sarai A, Nussinov R: **Protein-DNA interactions: Structural, thermodynamic and clustering patterns of conserved residues in DNA-binding proteins.** *Nucleic Acids Research* 2008, **36**(18):5922-5932.
16. HIV/AIDS JUNPo: **World aids day report 2011.** 2011.
17. Coffin JM: **Hiv population-dynamics in-vivo - implications for genetic-variation, pathogenesis, and therapy.** *Science* 1995, **267**(5197):483-489.
18. Smith C, Marks AD, Lieberman M: **Marks' basic medical biochemistry: A clinical approach, 2nd edition:** Lippincott Williams & Wilkins; 2004.
19. Huang HF, Chopra R, Verdine GL, Harrison SC: **Structure of a covalently trapped catalytic complex of hiv-i reverse transcriptase: Implications for drug resistance.** *Science* 1998, **282**(5394):1669-1675.
20. Jacobomolina A, Ding JP, Nanni RG, Clark AD, Lu XD, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P *et al*: **Crystal-structure of human-immunodeficiency-virus type-1 reverse-transcriptase complexed with double-stranded DNA at 3.0 angstrom resolution shows bent DNA.** *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(13):6320-6324.
21. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA: **Crystal-structure at 3.5 angstrom resolution of hiv-1 reverse-transcriptase complexed with an inhibitor.** *Science* 1992, **256**(5065):1783-1790.

Capítulo V: Referências

22. Tramontano E: **Hiv-1 rnase h: Recent progress in an exciting, yet little explored, drug target.** *Mini-Rev Med Chem* 2006, **6**(6):727-737.
23. Himmel DM, Sarafianos SG, Dharmasena S, Hossain MM, McCoy-Simandle K, Ilina T, Clark AD, Jr., Knight JL, Julias JG, Clark PK *et al*: **Hiv-1 reverse transcriptase structure with rnase h inhibitor dihydroxy benzoyl naphthyl hydrazone bound at a novel site.** *Acs Chemical Biology* 2006, **1**(11):702-712.
24. Davis WR, Tomsho J, Nikam S, Cook EM, Somand D, Peliska JA: **Inhibition of hiv-1 reverse transcriptase-catalyzed DNA strand transfer reactions by 4-chlorophenylhydrazone of mesoxalic acid.** *Biochemistry* 2000, **39**(46):14279-14291.
25. Menendez-Arias L: **Mechanisms of resistance to nucleoside analogue inhibitors of hiv-1 reverse transcriptase.** *Virus Research* 2008, **134**(1-2):124-146.
26. De Clercq E: **Antiviral drugs in current clinical use.** *J Clin Virol* 2004, **30**(2):115-133.
27. Peletskaya EN, Kogon AA, Tuske S, Arnold E, Hughes SH: **Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA.** *Journal of Virology* 2004, **78**(7):3387-3397.
28. Esnouf RM, Ren JS, Hopkins AL, Ross CK, Jones EY, Stammers DK, Stuart DI: **Unique features in the structure of the complex between hiv-1 reverse transcriptase and the bis(heteroaryl)piperazine (bhap) u-90152 explain resistance mutations for this nonnucleoside inhibitor.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(8):3984-3989.
29. De Clercq E, Holy A: **Acyclic nucleoside phosphonates: A key class of antiviral drugs.** *Nature Reviews Drug Discovery* 2005, **4**(11):928-940.
30. De Clercq E: **The acyclic nucleoside phosphonates from inception to clinical use: Historical perspective.** *Antiviral Research* 2007, **75**(1):1-13.
31. Balzarini J, Holy A, Jindrich J, Naesens L, Snoeck R, Schols D, Declercq E: **Differential antiherspesvirus and antiretrovirus effects of the (s) and (r) enantiomers of acyclic nucleoside phosphonates - potent and selective invitro and invivo antiretrovirus activities of (r)-9-(2-phosphonomethoxypropyl)-2,6-diaminopurine.** *Antimicrobial Agents and Chemotherapy* 1993, **37**(2):332-338.
32. Das K, Bandwar RP, White KL, Feng JY, Sarafianos SG, Tuske S, Tu X, Clark AD, Jr., Boyer PL, Hou X *et al*: **Structural basis for the role of the k65r mutation in hiv-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance.** *Journal of Biological Chemistry* 2009, **284**(50):35092-35100.
33. Margot NA, Waters JM, Miller MD: **In vitro human immunodeficiency virus type 1 resistance selections with combinations of tenofovir and emtricitabine or abacavir and lamivudine.** *Antimicrobial Agents and Chemotherapy* 2006, **50**(12):4087-4095.
34. Garcia-Lerma JG, MacInnes H, Bennett D, Reid P, Nidtha S, Weinstock H, Kaplan JE, Heneine W: **A novel genetic pathway of human immunodeficiency virus type 1 resistance to stavudine mediated by the k65r mutation.** *Journal of Virology* 2003, **77**(10):5685-5693.
35. Sousa SF, Fernandes PA, Ramos MJ: **Protein-ligand docking: Current status and future challenges.** *Proteins-Structure Function and Bioinformatics* 2006, **65**(1):15-26.
36. Kapetanovic IM: **Computer-aided drug discovery and development (cadd): In silico-chemico-biological approach.** *Chem-Biol Interact* 2008, **171**(2):165-176.
37. Sousa SF, Cerqueira NMFS, Fernandes PA, Ramos MJ: **Virtual screening in drug design and development.** *Combinatorial Chemistry & High Throughput Screening* 2010, **13**(5):442-453.
38. Allinger NL: **Conformational-analysis .130. Mm2 - hydrocarbon force-field utilizing v1 and v2 torsional terms.** *Journal of the American Chemical Society* 1977, **99**(25):8127-8134.
39. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **Charmm - a program for macromolecular energy, minimization, and dynamics calculations.** *Journal of Computational Chemistry* 1983, **4**(2):187-217.
40. Mackerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al*: **All-atom empirical potential for molecular modeling and dynamics studies of proteins.** *Journal of Physical Chemistry B* 1998, **102**(18):3586-3616.

41. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA: **A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules.** *Journal of the American Chemical Society* 1995, **117**(19):5179-5197.
42. Jorgensen WL, Maxwell DS, TiradoRives J: **Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids.** *Journal of the American Chemical Society* 1996, **118**(45):11225-11236.
43. Jorgensen WL, Tiradorives J: **The opls potential functions for proteins - energy minimizations for crystals of cyclic-peptides and crambin.** *Journal of the American Chemical Society* 1988, **110**(6):1657-1666.
44. Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT: **Derivation of class-ii force-fields .1. Methodology and quantum force-field for the alkyl functional-group and alkane molecules.** *Journal of Computational Chemistry* 1994, **15**(2):162-182.
45. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennel J, Torda AE, Huber T, Kruger P, van Gunsteren WF: **The gromos biomolecular simulation program package.** *J Phys Chem A* 1999, **103**(19):3596-3607.
46. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA: **Development and testing of a general amber force field.** *Journal of Computational Chemistry* 2004, **25**(9):1157-1174.
47. Alder BJ, Wainwright TE: **Studies in molecular dynamics .1. General method.** *Journal of Chemical Physics* 1959, **31**(2):459-466.
48. McCammon JA, Gelin BR, Karplus M: **Dynamics of folded proteins.** *Nature* 1977, **267**(5612):585-590.
49. Karplus M, McCammon JA: **Molecular dynamics simulations of biomolecules.** *Nature Structural Biology* 2002, **9**(9):646-652.
50. Swope WC, Andersen HC, Berens PH, Wilson KR: **A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules - application to small water clusters.** *Journal of Chemical Physics* 1982, **76**(1):637-649.
51. Ryckaert JP, Ciccotti G, Berendsen HJC: **Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes.** *Journal of Computational Physics* 1977, **23**(3):327-341.
52. Nose S: **A unified formulation of the constant temperature molecular-dynamics methods.** *Journal of Chemical Physics* 1984, **81**(1):511-519.
53. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR: **Molecular-dynamics with coupling to an external bath.** *Journal of Chemical Physics* 1984, **81**(8):3684-3690.
54. Loncharich RJ, Brooks BR, Pastor RW: **Langevin dynamics of peptides - the frictional dependence of isomerization rates of n-acetylalanyl-n'-methylamide.** *Biopolymers* 1992, **32**(5):523-535.
55. Izaguirre JA, Catarella DP, Wozniak JM, Skeel RD: **Langevin stabilization of molecular dynamics.** *Journal of Chemical Physics* 2001, **114**(5):2090-2098.
56. Feller SE, Zhang YH, Pastor RW, Brooks BR: **Constant-pressure molecular-dynamics simulation - the langevin piston method.** *Journal of Chemical Physics* 1995, **103**(11):4613-4621.
57. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, R.Luo, Merz KM, Pearlman DA, Crowley M *et al*: **Amber 9.** In. University of California, San Francisco; 2006.
58. Connolly ML: **Analytical molecular-surface calculation.** *Journal of Applied Crystallography* 1983, **16**(OCT):548-558.
59. Nicholls A, Sharp KA, Honig B: **Delphi.** In. Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY.; 1990.
60. Sharp KA, Honig B: **Calculating total electrostatic energies with the nonlinear poisson-boltzmann equation.** *Journal of Physical Chemistry* 1990, **94**(19):7684-7692.
61. Rocchia W, Alexov E, Honig B: **Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions.** *Journal of Physical Chemistry B* 2001, **105**(28):6507-6514.

Capítulo V: Referências

62. Chen SWW, Honig B: **Monovalent and divalent salt effects on electrostatic free energies defined by the nonlinear poisson-boltzmann equation: Application to DNA binding reactions.** *Journal of Physical Chemistry B* 1997, **101**(44):9113-9118.
63. Carra C, Saha J, Cucinotta FA: **Theoretical prediction of the binding free energy for mutants of replication protein a.** *J Mol Model* 2012, **18**(7):3035-3049.
64. Kumar MDS, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A: **Protherm and pronit: Thermodynamic databases for proteins and protein-nucleic acid interactions.** *Nucleic Acids Research* 2006, **34**:D204-D206.
65. Bas DC, Rogers DM, Jensen JH: **Very fast prediction and rationalization of pk(a) values for protein-ligand complexes.** *Proteins-Structure Function and Bioinformatics* 2008, **73**(3):765-783.
66. Li H, Robertson AD, Jensen JH: **Very fast empirical prediction and rationalization of protein pk(a) values.** *Proteins-Structure Function and Bioinformatics* 2005, **61**(4):704-721.
67. Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH: **Propka3: Consistent treatment of internal and surface residues in empirical pk(a) predictions.** *Journal of Chemical Theory and Computation* 2011, **7**(2):525-537.
68. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: **Comparison of simple potential functions for simulating liquid water.** *Journal of Chemical Physics* 1983, **79**(2):926-935.
69. Tan S, Richmond TJ: **Crystal structure of the yeast mat alpha 2/mcm1/DNA ternary complex.** *Nature* 1998, **391**(6668):660-666.
70. Schildbach JF, Karzai AW, Raumann BE, Sauer RT: **Origins of DNA-binding specificity: Role of protein contacts with the DNA backbone.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(3):811-817.
71. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y: **Solution structure of a specific DNA complex of the myb DNA-binding domain with cooperative recognition helices.** *Cell* 1994, **79**(4):639-648.
72. Murphy FV, Sweet RM, Churchill MEA: **The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition.** *Embo Journal* 1999, **18**(23):6610-6618.
73. Bochkarev A, Bochkareva E, Frappier L, Edwards AM: **The 2.2 angstrom structure of a permanganate-sensitive DNA site bound by the epstein-barr virus origin binding protein, ebna1.** *Journal of Molecular Biology* 1998, **284**(5):1273-1278.
74. Ribeiro JV, Cerqueira NMFSA, Moreira IS, Fernandes PA, Ramos MJ: **Compasm: An amber-vmd alanine scanning mutagenesis plug-in.** *Theoretical Chemistry Accounts - In press* 2012.
75. Brown BM, Milla ME, Smith TL, Sauer RT: **Scanning mutagenesis of the arc repressor as a functional probe of operator recognition.** *Nature Structural Biology* 1994, **1**(3):164-168.
76. Cruickshank J, Shire K, Davidson AR, Edwards AM, Frappier L: **Two domains of the epstein-barr virus origin DNA-binding protein, ebna1, orchestrate sequence-specific DNA binding.** *Journal of Biological Chemistry* 2000, **275**(29):22273-22277.
77. Oda M, Furukawa K, Ogata K, Sarai A, Ishii S, Nishimura Y, Nakamura H: **Identification of indispensable residues for specific DNA-binding in the imperfect tandem repeats of c-myb r2r3.** *Protein Engineering* 1997, **10**(12):1407-1414.
78. Oda M, Furukawa K, Ogata K, Sarai A, Nakamura H: **Thermodynamics of specific and non-specific DNA binding by the c-myb DNA-binding domain.** *Journal of Molecular Biology* 1998, **276**(3):571-590.
79. Klass J, Murphy FV, Fouts S, Serenil M, Changela A, Siple J, Churchill MEA: **The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity.** *Nucleic Acids Research* 2003, **31**(11):2852-2864.
80. Acton TB, Mead J, Steiner AM, Vershon AK: **Scanning mutagenesis of mcm1: Residues required for DNA binding, DNA bending, and transcriptional activation by a mads-box protein.** *Molecular and Cellular Biology* 2000, **20**(1):1-11.

81. Wang JM, Wang W, Kollman PA, Case DA: **Automatic atom type and bond type perception in molecular mechanical calculations.** *J Mol Graph* 2006, **25**(2):247-260.
82. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA *et al*: **Gaussian 09, revision a.1.** *Gaussian, Inc, Wallingford CT* 2009.
83. Tuske S, Sarafianos SG, Clark AD, Ding JP, Naeger LK, White KL, Miller MD, Gibbs CS, Boyer PL, Clark P *et al*: **Structures of hiv-1 rt-DNA complexes before and after incorporation of the anti-aids drug tenofovir.** *Nature Structural & Molecular Biology* 2004, **11**(5):469-474.
84. Sarafianos SG, Clark AD, Das K, Tuske S, Birktoft JJ, Ilankumaran P, Ramesha AR, Sayer JM, Jerina DM, Boyer PL *et al*: **Structures of hiv-1 reverse transcriptase with pre- and post-translocation aztmp-terminated DNA.** *Embo Journal* 2002, **21**(23):6614-6624.
85. Humphrey W, Dalke A, Schulten K: **Vmd: Visual molecular dynamics.** *Journal of Molecular Graphics* 1996, **14**(1):33-&.
86. Cerqueira NMFSA, Ribeiro J, Fernandes PA, Ramos MJ: **Vslab-an implementation for virtual high-throughput screening using autodock and vmd.** *International Journal of Quantum Chemistry* 2011, **111**(6):1208-1212.
87. Misra VK, Hecht JL, Sharp KA, Friedman RA, Honig B: **Salt effects on protein-DNA interactions - the lambda-ci repressor and ecori endonuclease.** *Journal of Molecular Biology* 1994, **238**(2):264-280.
88. Misra VK, Sharp KA, Friedman RA, Honig B: **Salt effects on ligand-DNA binding - minor-groove binding antibiotics.** *Journal of Molecular Biology* 1994, **238**(2):245-263.
89. Sheinerman FB, Honig B: **On the role of electrostatic interactions in the design of protein-protein interfaces.** *Journal of Molecular Biology* 2002, **318**(1):161-177.
90. Fernandez-Recio J: **Prediction of protein binding sites and hot spots.** *Wiley Interdisciplinary Reviews-Computational Molecular Science* 2011, **1**(5):680-698.
91. Guharoy M, Chakrabarti P: **Conservation and relative importance of residues across protein-protein interfaces.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15447-15452.
92. Moreira IS, Fernandes PA, Ramos MJ: **Hot spot occlusion from bulk water: A comprehensive study of the complex between the lysozyme hel and the antibody fvd1.3.** *Journal of Physical Chemistry B* 2007, **111**(10):2697-2706.
93. Marchand B, White KL, Ly JK, Margot NA, Wang R, McDermott M, Miller MD, Gotte M: **Effects of the translocation status of human immunodeficiency virus type 1 reverse transcriptase on the efficiency of excision of tenofovir.** *Antimicrobial Agents and Chemotherapy* 2007, **51**(8):2911-2919.
94. Bertrand HO, Ha-Duong T, Fermandjian S, Hartmann B: **Flexibility of the b-DNA backbone: Effects of local and neighbouring sequences on pyrimidine-purine steps.** *Nucleic Acids Research* 1998, **26**(5):1261-1267.

Capítulo V: Referências

Capítulo VI. Informação Suplementar

Tabela VI.1 – Resultados da energia livre de ligação (kcal mol⁻¹) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns e 6 ns em solvente explícito. Constantes dielétricas de 1 a 10.

Proteína	AA Mutado	$\Delta\Delta G_{exp}$ / kcal mol ⁻¹	$\Delta\Delta G_{MM-PBSA}$ / kcal mol ⁻¹																			
			2 ns										6 ns									
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
1BDT	M	1,10	3,72	4,09	4,68	4,79	4,85	4,89	4,92	4,93	4,87	4,96	2,84	1,72	4,31	4,48	4,58	4,64	4,69	4,72	4,40	4,76
1BDT	S	1,30	7,41	3,61	2,18	1,52	1,12	0,86	0,67	0,52	0,36	0,32	7,68	3,38	2,41	1,75	1,36	1,09	0,90	0,75	0,53	0,55
1BDT	K	1,30	221,81	251,29	76,46	58,29	47,40	40,13	34,94	31,05	56,03	25,61	223,65	113,59	77,09	58,77	47,78	40,46	35,22	31,30	28,19	25,81
1BDT	M	1,90	-1,40	-1,74	-0,36	-0,23	-0,15	-0,10	-0,06	-0,03	-0,21	0,01	-1,16	-0,48	-0,26	-0,15	-0,08	-0,04	-0,01	0,02	0,06	0,05
1BDT	Q	1,80	5,07	1,75	4,42	4,33	4,28	4,25	4,22	4,20	3,69	4,17	6,16	2,76	4,77	4,59	4,48	4,41	4,36	4,31	3,83	4,26
1BDT	N	2,00	3,85	2,59	3,40	3,35	3,32	3,29	3,28	3,27	3,08	3,26	5,29	3,34	3,90	3,74	3,64	3,57	3,53	3,50	3,30	3,45
1BDT	R	6,00	319,29	365,77	112,64	86,73	71,16	60,76	53,32	47,73	83,29	39,89	327,19	168,32	116,26	89,80	73,90	63,27	55,66	49,94	45,21	41,91
1BDT	R	0,20	276,42	317,31	95,61	72,98	59,39	50,32	43,84	38,97	70,19	32,14	277,79	140,21	96,08	73,34	59,68	50,57	44,05	39,16	34,94	32,29
1BDT	N	-1,00	4,14	4,73	1,42	1,08	0,88	0,75	0,66	0,58	1,04	0,49	3,68	1,95	1,28	0,98	0,81	0,69	0,60	0,54	0,48	0,45
1BDT	R	-0,30	207,84	232,37	71,15	54,08	43,85	37,03	32,17	28,52	51,56	23,43	207,50	105,15	71,04	54,00	43,78	36,97	32,12	28,48	25,62	23,39
1BDT	S	-2,30	12,41	7,57	4,14	3,10	2,48	2,06	1,77	1,55	1,63	1,24	13,21	5,79	4,57	3,50	2,85	2,43	2,12	1,90	1,59	1,58
1BDT	V	-2,30	9,92	4,57	3,71	2,93	2,47	2,16	1,93	1,77	1,48	1,53	9,65	4,93	3,70	2,95	2,51	2,21	1,99	1,83	1,63	1,61
1BDT	N	3,20	-4,31	-4,03	-1,08	-0,68	-0,44	-0,27	-0,16	-0,07	-0,37	0,05	-4,59	-3,66	-1,13	-0,70	-0,45	-0,28	-0,16	-0,07	-0,27	0,06
1BDT	S	-0,20	2,08	-0,29	0,90	0,75	0,67	0,62	0,58	0,54	0,29	0,51	2,63	0,84	1,06	0,87	0,76	0,68	0,62	0,58	0,45	0,52
1BDT	Q	-0,40	0,17	-0,36	0,08	0,06	0,06	0,05	0,05	0,04	-0,08	0,04	-0,45	-0,30	-0,16	-0,13	-0,10	-0,09	-0,08	-0,07	-0,07	-0,06
1B3T	R	3,41	299,89	342,21	110,52	87,17	73,19	63,87	57,20	52,18	84,56	45,12	303,49	342,89	111,64	87,98	73,81	64,36	57,60	52,52	84,72	45,37
1B3T	Y	2,62	11,22	-1,84	8,08	7,70	7,46	7,29	7,15	7,04	5,82	6,88	11,44	-2,70	7,46	6,96	6,65	6,43	6,27	6,14	4,79	5,94
1B3T	R	4,40	298,46	315,73	105,60	81,46	66,95	57,24	50,29	45,05	72,46	37,68	298,57	315,51	105,61	81,43	66,88	57,15	50,17	44,93	72,28	37,54
1MSE	S	0,06	3,37	2,55	1,23	0,96	0,80	0,69	0,61	0,55	0,60	0,47	6,91	4,38	2,21	1,62	1,26	1,02	0,85	0,72	0,77	0,54
1MSE	N	0,60	1,03	2,09	0,42	0,34	0,29	0,26	0,23	0,22	0,47	0,19	0,79	2,23	0,35	0,30	0,26	0,24	0,22	0,21	0,49	0,19
1MSE	E	-0,10	-121,75	-139,63	-41,68	-31,68	-25,69	-21,69	-18,84	-16,70	-30,90	-13,72	-126,01	-144,26	-43,14	-32,79	-26,58	-22,45	-19,50	-17,29	-31,94	-14,20
1MSE	S	0,10	5,52	3,27	2,42	2,02	1,79	1,63	1,52	1,43	1,35	1,31	5,59	2,38	2,22	1,80	1,54	1,37	1,25	1,16	0,94	1,02
1QRV	L	0,02	3,71	3,59	3,85	3,86	3,88	3,89	3,89	3,90	3,91	3,91	2,73	2,35	3,24	3,30	3,34	3,36	3,38	3,39	3,28	3,41
1QRV	M	1,20	9,10	8,80	8,78	8,73	8,70	8,68	8,67	8,66	8,66	8,64	9,16	6,53	8,50	8,41	8,36	8,32	8,30	8,28	7,86	8,25
1QRV	V	-0,30	2,43	2,34	3,59	3,73	3,82	3,88	3,92	3,96	3,88	4,00	2,29	2,13	3,87	4,07	4,19	4,27	4,32	4,37	4,17	4,43
1MNM	K	0,54	347,93	392,94	122,31	94,09	77,15	65,84	57,75	51,68	89,67	43,17	322,90	163,82	112,84	86,56	70,78	60,25	52,73			
1MNM	E	0,24	-278,00	-322,00	-95,69	-72,90	-59,23	-50,13	-43,63	-38,75	-71,11	-31,91	-286,38	-145,64	-98,47	-75,00	-60,91	-51,52	-44,81			

Capítulo VI: Informação Suplementar

1MNM	R	0,28	313,30	354,47	109,22	83,67	68,34	58,10	50,78	45,28	79,67	37,57	321,17	162,79	112,13	85,97	70,26	59,78	52,28
1MNM	R	0,81	385,14	439,47	143,26	113,02	94,88	82,77	74,12	67,62	108,44	58,50	386,72	200,49	144,14	113,81	95,61	83,46	74,77
1MNM	K	0,64	346,36	400,27	121,95	93,85	76,98	65,71	57,66	51,60	90,65	43,10	348,32	176,85	122,33	94,04	77,05	65,71	57,59
1MNM	I	0,88	-1,12	0,87	0,98	1,24	1,39	1,49	1,55	1,60	1,62	1,67	-1,08	0,41	0,98	1,23	1,38	1,47	1,54
1MNM	E	0,20	-287,02	-330,70	-99,02	-75,53	-61,43	-52,04	-45,32	-40,29	-73,25	-33,23	-271,12	-137,71	-93,44	-71,25	-57,93	-49,05	-42,71
1MNM	I	0,86	4,60	2,40	1,97	1,64	1,43	1,30	1,20	1,12	0,94	1,01	3,62	2,16	1,60	1,34	1,18	1,07	0,99
1MNM	K	-0,13	243,80	282,74	84,02	64,06	52,08	44,10	38,40	34,12	62,71	28,14	247,50	125,82	85,38	65,12	52,96	44,86	39,07
1MNM	F	4,05	1,34	0,18	0,44	0,33	0,27	0,23	0,20	0,17	0,05	0,14	0,07	-0,09	0,04	0,03	0,03	0,04	0,03
1MNM	I	4,05	2,71	2,04	1,21	1,01	0,90	0,82	0,77	0,72	0,74	0,66	2,95	1,70	1,27	1,06	0,94	0,85	0,79
1MNM	E	0,41	-218,57	-256,99	-75,20	-57,29	-46,55	-39,40	-34,29	-30,46	-56,99	-25,10	-221,74	-112,88	-76,24	-58,08	-47,18	-39,92	-34,74
1MNM	N	1,75	-1,32	-3,35	-0,39	-0,27	-0,20	-0,16	-0,13	-0,10	-0,61	-0,07	-2,91	-1,43	-0,86	-0,61	-0,46	-0,36	-0,28
1MNM	K	-0,13	225,68	258,31	77,57	59,07	47,98	40,58	35,31	31,35	57,37	25,82	234,29	119,00	80,69	61,49	49,98	42,31	36,83
1MNM	R	2,76	246,41	283,71	84,59	64,40	52,30	44,23	38,49	34,18	62,91	28,16	248,29	125,89	85,29	64,95	52,76	44,64	38,85
1MNM	H	-0,81	4,05	1,63	1,76	1,48	1,31	1,19	1,11	1,05	0,85	0,96	3,16	1,97	1,70	1,51	1,40	1,33	1,28
1MNM	V	0,58	-3,27	-0,88	-0,31	0,06	0,29	0,43	0,54	0,62	0,63	0,73	-3,28	-0,21	0,14	0,56	0,82	0,99	1,10
1MNM	T	0,64	-3,02	-4,54	-0,66	-0,36	-0,17	-0,05	0,04	0,11	-0,37	0,21	-2,84	-1,47	-0,58	-0,30	-0,12	-0,01	0,08
1MNM	F	4,05	1,31	0,20	0,47	0,37	0,31	0,27	0,24	0,21	0,09	0,18	1,08	0,50	0,40	0,31	0,26	0,23	0,21
1MNM	S	0,43	-1,12	1,02	-0,30	-0,19	-0,13	-0,09	-0,06	-0,04	0,23	-0,01	-1,39	-0,70	-0,38	-0,26	-0,18	-0,13	-0,10
1MNM	K	4,05	347,57	411,44	123,35	95,29	78,43	67,17	59,11	53,06	93,45	44,56	350,61	179,65	124,94	96,69	79,73	68,40	60,29
1MNM	R	4,05	389,81	429,99	139,10	107,66	88,75	76,10	67,04	60,22	98,76	50,62	391,55	195,31	139,33	107,70	88,68	75,95	66,83
1MNM	K	2,83	267,91	307,32	92,30	70,36	57,21	48,43	42,17	37,47	68,20	30,91	270,73	137,19	93,34	71,18	57,88	49,02	42,69
1MNM	H	-1,24	5,14	1,62	2,11	1,72	1,48	1,31	1,19	1,10	0,68	0,96	5,30	2,94	2,22	1,82	1,58	1,41	1,29
1MNM	I	3,24	4,17	1,47	1,49	1,16	0,96	0,82	0,73	0,65	0,47	0,55	3,76	1,86	1,36	1,05	0,88	0,76	0,67
1MNM	M	4,05	1,05	0,11	0,41	0,34	0,28	0,25	0,23	0,21	0,10	0,18	0,79	0,40	0,31	0,25	0,21	0,19	0,17
1MNM	K	4,05	359,02	407,64	125,53	96,30	78,73	67,00	58,61	52,30	91,34	43,45	361,43	181,83	126,36	96,92	79,25	67,44	58,99
1MNM	K	4,05	415,10	445,56	145,90	112,26	92,06	78,57	68,94	61,71	101,15	51,56	403,58	200,22	141,87	109,17	89,54	76,44	67,08
1MNM	F	-0,13	0,77	0,56	0,34	0,29	0,25	0,23	0,21	0,20	0,11	0,18	-0,34	0,04	0,00	0,04	0,06	0,07	0,08
1MNM	E	2,27	-357,56	-407,99	-123,08	-93,78	-76,20	-64,48	-56,10	-49,81	-89,04	-40,99	-359,77	-182,02	-123,63	-94,14	-76,45	-64,66	-56,24
1MNM	S	0,00	2,64	2,93	0,98	0,77	0,65	0,56	0,50	0,45	0,66	0,38	2,65	1,23	0,98	0,76	0,63	0,55	0,49
1MNM	V	0,11	6,38	3,21	2,10	1,57	1,26	1,04	0,89	0,78	0,71	0,62	5,26	2,53	1,76	1,32	1,06	0,89	0,76
1MNM	L	4,05	-0,69	-1,58	0,28	0,40	0,48	0,52	0,55	0,57	0,19	0,60	-1,61	-0,87	0,10	0,31	0,43	0,51	0,57
1MNM	T	4,05	4,41	3,19	1,54	1,19	0,97	0,83	0,73	0,65	0,75	0,54	3,20	1,67	1,15	0,90	0,74	0,64	0,57
1MNM	T	0,64	2,24	0,23	0,71	0,52	0,41	0,33	0,28	0,24	0,05	0,18	1,27	0,71	0,44	0,33	0,27	0,23	0,20

Tabela VI.2 – Resultados da energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito e em solvente implícito. Constantes dielétricas de 1 a 10.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$																			
			Solvente Explícito										Solvente Implícito									
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
1BDT	M	1,10	3,72	4,09	4,68	4,79	4,85	4,89	4,92	4,93	4,87	4,96	1,24	0,16	1,38	1,39	1,39	1,39	1,39	1,38	1,29	1,37
1BDT	S	1,30	7,41	3,61	2,18	1,52	1,12	0,86	0,67	0,52	0,36	0,32	11,96	3,66	2,85	1,70	1,00	0,54	0,20	-0,06	-0,52	-0,43
1BDT	K	1,30	221,81	251,29	76,46	58,29	47,40	40,13	34,94	31,05	56,03	25,61	270,91	132,84	94,76	72,68	59,40	50,53	44,18	39,41	34,39	32,71
1BDT	M	1,90	-1,40	-1,74	-0,36	-0,23	-0,15	-0,10	-0,06	-0,03	-0,21	0,01	1,80	0,89	0,69	0,55	0,47	0,41	0,37	0,34	0,31	0,30
1BDT	Q	1,80	5,07	1,75	4,42	4,33	4,28	4,25	4,22	4,20	3,69	4,17	1,86	1,65	4,01	4,28	4,44	4,55	4,63	4,69	4,38	4,33
1BDT	N	2,00	3,85	2,59	3,40	3,35	3,32	3,29	3,28	3,27	3,08	3,26	5,43	3,84	4,00	3,82	3,72	3,65	3,60	3,56	3,43	3,51
1BDT	R	6,00	319,29	365,77	112,64	86,73	71,16	60,76	53,32	47,73	83,29	39,89	345,34	175,80	124,66	96,89	80,17	68,98	60,96	54,92	49,31	46,43
1BDT	R	0,20	276,42	317,31	95,61	72,98	59,39	50,32	43,84	38,97	70,19	32,14	271,53	137,64	96,42	74,48	61,30	52,50	46,20	41,47	37,22	34,82
1BDT	N	-1,00	4,14	4,73	1,42	1,08	0,88	0,75	0,66	0,58	1,04	0,49	3,81	1,84	1,31	1,00	0,81	0,69	0,61	0,54	0,46	0,45
1BDT	R	-0,30	207,84	232,37	71,15	54,08	43,85	37,03	32,17	28,52	51,56	23,43	210,78	106,39	72,07	54,76	44,38	37,46	32,53	28,84	25,85	23,67
1BDT	S	-2,30	12,41	7,57	4,14	3,10	2,48	2,06	1,77	1,55	1,63	1,24	17,56	9,10	5,32	3,78	2,85	2,23	1,78	1,44	1,24	0,96
1BDT	V	-2,30	9,92	4,57	3,71	2,93	2,47	2,16	1,93	1,77	1,48	1,53	10,22	4,64	4,01	3,23	2,77	2,46	2,24	2,08	1,78	1,85
1BDT	N	3,20	-4,31	-4,03	-1,08	-0,68	-0,44	-0,27	-0,16	-0,07	-0,37	0,05	-1,03	-0,35	0,36	0,52	0,61	0,67	0,71	0,73	0,73	0,76
1BDT	S	-0,20	2,08	-0,29	0,90	0,75	0,67	0,62	0,58	0,54	0,29	0,51	5,69	2,63	2,16	1,72	1,45	1,27	1,14	1,04	0,92	0,90
1BDT	Q	-0,40	0,17	-0,36	0,08	0,06	0,06	0,05	0,05	0,04	-0,08	0,04	-1,74	-0,77	-0,57	-0,42	-0,33	-0,27	-0,23	-0,20	-0,17	-0,16
1B3T	R	3,41	299,89	342,21	110,52	87,17	73,19	63,87	57,20	52,18	84,56	45,12	317,57	369,59	118,15	93,38	78,45	68,44	61,24	55,81	90,42	48,11
1B3T	Y	2,62	11,22	-1,84	8,08	7,70	7,46	7,29	7,15	7,04	5,82	6,88	-1,57	-8,38	5,51	6,29	6,69	6,91	7,04	7,12	6,11	7,20
1B3T	R	4,40	298,46	315,73	105,60	81,46	66,95	57,24	50,29	45,05	72,46	37,68	295,68	311,66	105,00	81,20	66,87	57,28	50,39	45,20	72,25	37,86
1MSE	S	0,06	3,37	2,55	1,23	0,96	0,80	0,69	0,61	0,55	0,60	0,47	13,46	6,20	4,85	3,76	3,10	2,65	2,32	2,08	1,73	1,72
1MSE	N	0,60	1,03	2,09	0,42	0,34	0,29	0,26	0,23	0,22	0,47	0,19	0,68	0,31	0,32	0,28	0,25	0,23	0,22	0,21	0,17	0,19
1MSE	E	-0,10	-121,75	-139,63	-41,68	-31,68	-25,69	-21,69	-18,84	-16,70	-30,90	-13,72	-129,18	-64,99	-44,18	-33,57	-27,21	-22,97	-19,94	-17,68	-15,82	-14,52
1MSE	S	0,10	5,52	3,27	2,42	2,02	1,79	1,63	1,52	1,43	1,35	1,31	10,99	5,68	4,02	3,15	2,62	2,27	2,01	1,82	1,66	1,56
1QRV	L	0,02	3,71	3,59	3,85	3,86	3,88	3,89	3,89	3,90	3,91	3,91	2,57	1,33	2,11	2,05	2,01	1,99	1,97	1,96	1,79	1,94
1QRV	M	1,20	9,10	8,80	8,78	8,73	8,70	8,68	8,67	8,66	8,66	8,64	3,27	4,08	5,33	5,58	5,72	5,82	5,89	5,94	5,96	6,01
1QRV	V	-0,30	2,43	2,34	3,59	3,73	3,82	3,88	3,92	3,96	3,88	4,00	12,71	5,53	5,73	4,86	4,34	3,99	3,74	3,55	3,11	3,29
1MNM	K	0,54	347,93	392,94	122,31	94,09	77,15	65,84	57,75	51,68	89,67	43,17	337,52	383,29	119,17	91,88	75,49	64,55	56,73	50,86	88,08	42,62
1MNM	E	0,24	-278,00	-322,00	-95,69	-72,90	-59,23	-50,13	-43,63	-38,75	-71,11	-31,91	-325,23	-368,25	-108,98	-81,99	-65,82	-55,05	-47,36	-41,59	-78,01	-33,53

Capítulo VI: Informação Suplementar

1MNM	R	0,28	313,30	354,47	109,22	83,67	68,34	58,10	50,78	45,28	79,67	37,57	329,01	367,16	114,72	87,90	71,78	61,01	53,31	47,52	82,57	39,40
1MNM	R	0,81	385,14	439,47	143,26	113,02	94,88	82,77	74,12	67,62	108,44	58,50	376,14	422,79	140,05	110,52	92,77	80,92	72,45	66,08	105,14	57,13
1MNM	K	0,64	346,36	400,27	121,95	93,85	76,98	65,71	57,66	51,60	90,65	43,10	364,95	412,68	127,79	98,10	80,26	68,34	59,81	53,41	92,81	44,40
1MNM	I	0,88	-1,12	0,87	0,98	1,24	1,39	1,49	1,55	1,60	1,62	1,67	1,06	2,13	1,55	1,61	1,63	1,64	1,64	1,64	1,60	1,64
1MNM	E	0,20	-287,02	-330,70	-99,02	-75,53	-61,43	-52,04	-45,32	-40,29	-73,25	-33,23	-248,55	-289,92	-85,62	-65,26	-53,04	-44,91	-39,10	-34,74	-64,26	-28,65
1MNM	I	0,86	4,60	2,40	1,97	1,64	1,43	1,30	1,20	1,12	0,94	1,01	2,64	1,30	0,99	0,78	0,66	0,57	0,50	0,45	0,25	0,38
1MNM	K	-0,13	243,80	282,74	84,02	64,06	52,08	44,10	38,40	34,12	62,71	28,14	248,15	288,74	85,67	65,36	53,18	45,05	39,25	34,90	64,06	28,80
1MNM	F	4,05	1,34	0,18	0,44	0,33	0,27	0,23	0,20	0,17	0,05	0,14	0,35	0,63	0,15	0,13	0,12	0,10	0,10	0,10	0,10	0,09
1MNM	I	4,05	2,71	2,04	1,21	1,01	0,90	0,82	0,77	0,72	0,74	0,66	1,84	1,52	1,10	1,01	0,95	0,89	0,86	0,83	0,73	0,78
1MNM	E	0,41	-218,57	-256,99	-75,20	-57,29	-46,55	-39,40	-34,29	-30,46	-56,99	-25,10	-222,75	-260,46	-76,43	-58,16	-47,21	-39,92	-34,72	-30,81	-57,73	-25,37
1MNM	N	1,75	-1,32	-3,35	-0,39	-0,27	-0,20	-0,16	-0,13	-0,10	-0,61	-0,07	0,11	-1,93	0,29	0,32	0,33	0,34	0,35	0,35	-0,08	0,35
1MNM	K	-0,13	225,68	258,31	77,57	59,07	47,98	40,58	35,31	31,35	57,37	25,82	314,84	342,37	109,56	83,90	68,50	58,22	50,88	45,37	77,20	37,64
1MNM	R	2,76	246,41	283,71	84,59	64,40	52,30	44,23	38,49	34,18	62,91	28,16	260,48	299,20	89,47	68,14	55,36	46,84	40,76	36,21	66,31	29,85
1MNM	H	-0,81	4,05	1,63	1,76	1,48	1,31	1,19	1,11	1,05	0,85	0,96	4,53	0,82	3,36	3,22	3,14	3,08	3,04	3,01	2,48	2,97
1MNM	V	0,58	-3,27	-0,88	-0,31	0,06	0,29	0,43	0,54	0,62	0,63	0,73	-4,67	0,11	1,16	1,87	2,30	2,57	2,77	2,91	3,06	3,11
1MNM	T	0,64	-3,02	-4,54	-0,66	-0,36	-0,17	-0,05	0,04	0,11	-0,37	0,21	7,69	3,31	2,76	2,17	1,81	1,57	1,40	1,27	1,06	1,10
1MNM	F	4,05	1,31	0,20	0,47	0,37	0,31	0,27	0,24	0,21	0,09	0,18	0,15	-0,27	0,11	0,12	0,12	0,12	0,12	0,13	0,03	0,13
1MNM	S	0,43	-1,12	1,02	-0,30	-0,19	-0,13	-0,09	-0,06	-0,04	0,23	-0,01	-0,65	1,28	-0,04	0,03	0,08	0,10	0,12	0,13	0,37	0,15
1MNM	K	4,05	347,57	411,44	123,35	95,29	78,43	67,17	59,11	53,06	93,45	44,56	387,64	425,85	137,19	105,91	87,12	74,57	65,59	58,84	97,83	49,35
1MNM	R	4,05	389,81	429,99	139,10	107,66	88,75	76,10	67,04	60,22	98,76	50,62	396,47	423,44	140,75	108,72	89,45	76,55	67,32	60,37	97,22	50,58
1MNM	K	2,83	267,91	307,32	92,30	70,36	57,21	48,43	42,17	37,47	68,20	30,91	273,96	313,79	94,59	72,18	58,72	49,74	43,34	38,53	69,64	31,79
1MNM	H	-1,24	5,14	1,62	2,11	1,72	1,48	1,31	1,19	1,10	0,68	0,96	1,53	-4,86	1,06	1,00	0,96	0,93	0,91	0,89	-0,30	0,85
1MNM	I	3,24	4,17	1,47	1,49	1,16	0,96	0,82	0,73	0,65	0,47	0,55	3,62	0,03	1,27	0,98	0,81	0,69	0,60	0,55	0,21	0,46
1MNM	M	4,05	1,05	0,11	0,41	0,34	0,28	0,25	0,23	0,21	0,10	0,18	2,70	1,17	0,94	0,73	0,59	0,50	0,44	0,39	0,31	0,32
1MNM	K	4,05	359,02	407,64	125,53	96,30	78,73	67,00	58,61	52,30	91,34	43,45	383,88	421,61	134,64	103,47	84,73	72,21	63,26	56,52	94,65	47,05
1MNM	K	4,05	415,10	445,56	145,90	112,26	92,06	78,57	68,94	61,71	101,15	51,56	429,52	435,75	148,41	113,39	92,39	78,39	68,39	60,90	97,50	50,38
1MNM	F	-0,13	0,77	0,56	0,34	0,29	0,25	0,23	0,21	0,20	0,11	0,18	-0,06	0,26	0,08	0,10	0,12	0,12	0,12	0,12	0,03	0,12
1MNM	E	2,27	-357,56	-407,99	-123,08	-93,78	-76,20	-64,48	-56,10	-49,81	-89,04	-40,99	-362,27	-408,66	-124,21	-94,52	-76,73	-64,88	-56,42	-50,07	-88,92	-41,19
1MNM	S	0	2,64	2,93	0,98	0,77	0,65	0,56	0,50	0,45	0,66	0,38	-0,30	0,00	-0,11	-0,08	-0,06	-0,06	-0,05	-0,04	-0,02	-0,03
1MNM	V	0,11	6,38	3,21	2,10	1,57	1,26	1,04	0,89	0,78	0,71	0,62	5,26	3,43	1,77	1,35	1,08	0,90	0,78	0,69	0,72	0,55
1MNM	L	4,05	-0,69	-1,58	0,28	0,40	0,48	0,52	0,55	0,57	0,19	0,60	1,06	-3,85	1,56	1,60	1,62	1,61	1,61	1,60	0,59	1,57
1MNM	T	4,05	4,41	3,19	1,54	1,19	0,97	0,83	0,73	0,65	0,75	0,54	3,78	3,16	1,39	1,10	0,92	0,80	0,72	0,65	0,80	0,55
1MNM	T	0,64	2,24	0,23	0,71	0,52	0,41	0,33	0,28	0,24	0,05	0,18	1,42	0,91	0,53	0,42	0,36	0,30	0,27	0,25	0,22	0,21

Tabela VI.3 – Resultados da energia livre de ligação (kcal mol^{-1}) calculadas pelo método ASM usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, e utilizando outra alternativa para o cálculo da energia livre de ligação, utilizando duas trajetórias. Constantes dielétricas de 1 a 10.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$																			
			2 ns										Monómeros									
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
1BDT	M	1,10	3,72	4,09	4,68	4,79	4,85	4,89	4,92	4,93	4,87	4,96	0,43	3,73	4,10	4,54	4,81	4,98	5,10	5,19	5,30	5,31
1BDT	S	1,30	7,41	3,61	2,18	1,52	1,12	0,86	0,67	0,52	0,36	0,32	5,99	3,20	1,23	0,64	0,29	0,05	-0,12	-0,24	-0,24	-0,42
1BDT	K	1,30	221,81	251,29	76,46	58,29	47,40	40,13	34,94	31,05	56,03	25,61	222,01	253,92	76,66	58,50	47,61	40,34	35,16	31,27	56,77	25,83
1BDT	M	1,90	-1,40	-1,74	-0,36	-0,23	-0,15	-0,10	-0,06	-0,03	-0,21	0,01	0,02	-2,09	0,23	0,25	0,27	0,28	0,29	0,29	-0,13	0,30
1BDT	Q	1,80	5,07	1,75	4,42	4,33	4,28	4,25	4,22	4,20	3,69	4,17	3,68	0,90	2,56	2,41	2,33	2,27	2,22	2,19	1,85	2,14
1BDT	N	2,00	3,85	2,59	3,40	3,35	3,32	3,29	3,28	3,27	3,08	3,26	3,84	1,85	2,35	2,16	2,05	1,98	1,93	1,89	1,69	1,83
1BDT	R	6,00	319,29	365,77	112,64	86,73	71,16	60,76	53,32	47,73	83,29	39,89	329,48	366,88	113,88	86,90	70,71	59,91	52,20	46,41	81,48	38,29
1BDT	R	0,20	276,42	317,31	95,61	72,98	59,39	50,32	43,84	38,97	70,19	32,14	274,99	316,18	94,55	71,98	58,42	49,38	42,91	38,06	69,39	31,25
1BDT	N	-1,00	4,14	4,73	1,42	1,08	0,88	0,75	0,66	0,58	1,04	0,49	4,98	5,17	1,88	1,49	1,27	1,11	1,00	0,92	1,40	0,81
1BDT	R	-0,30	207,84	232,37	71,15	54,08	43,85	37,03	32,17	28,52	51,56	23,43	206,77	229,06	70,60	53,59	43,40	36,61	31,76	28,13	50,68	23,05
1BDT	S	-2,30	12,41	7,57	4,14	3,10	2,48	2,06	1,77	1,55	1,63	1,24	8,91	4,68	2,62	1,83	1,36	1,04	0,82	0,65	0,63	0,41
1BDT	V	-2,30	9,92	4,57	3,71	2,93	2,47	2,16	1,93	1,77	1,48	1,53	8,36	3,44	3,01	2,34	1,94	1,67	1,48	1,34	1,06	1,14
1BDT	N	3,20	-4,31	-4,03	-1,08	-0,68	-0,44	-0,27	-0,16	-0,07	-0,37	0,05	-3,27	-3,80	-0,33	0,03	0,25	0,39	0,50	0,57	0,16	0,68
1BDT	S	-0,20	2,08	-0,29	0,90	0,75	0,67	0,62	0,58	0,54	0,29	0,51	-0,43	-1,05	0,55	0,68	0,75	0,80	0,83	0,86	0,65	0,89
1BDT	Q	-0,40	0,17	-0,36	0,08	0,06	0,06	0,05	0,05	0,04	-0,08	0,04	-1,04	0,18	-0,32	-0,23	-0,18	-0,14	-0,11	-0,09	0,04	-0,06
1B3T	R	3,41	299,89	342,21	110,52	87,17	73,19	63,87	57,20	52,18	84,56	45,12	699,19	341,02	109,49	85,70	71,50	62,07	55,34	50,30	82,71	43,24
1B3T	Y	2,62	11,22	-1,84	8,08	7,70	7,46	7,29	7,15	7,04	5,82	6,88	-13,20	-0,54	9,21	9,00	8,88	8,80	8,74	8,69	7,62	8,62
1B3T	R	4,40	298,46	315,73	105,60	81,46	66,95	57,24	50,29	45,05	72,46	37,68	628,77	315,28	104,47	80,27	65,76	56,10	49,20	44,03	71,86	36,79
1MSE	S	0,06	3,37	2,55	1,23	0,96	0,80	0,69	0,61	0,55	0,60	0,47	0,94	2,08	0,40	0,34	0,30	0,27	0,26	0,25	0,49	0,23
1MSE	N	0,60	1,03	2,09	0,42	0,34	0,29	0,26	0,23	0,22	0,47	0,19	8,19	6,56	4,18	3,68	3,38	3,18	3,04	2,94	3,11	2,79
1MSE	E	-0,10	-121,75	-139,63	-41,68	-31,68	-25,69	-21,69	-18,84	-16,70	-30,90	-13,72	-121,70	-134,01	-42,92	-33,07	-27,16	-23,22	-20,40	-18,29	-31,20	-15,34
1MSE	S	0,10	5,52	3,27	2,42	2,02	1,79	1,63	1,52	1,43	1,35	1,31	3,14	2,55	1,51	1,30	1,17	1,09	1,02	0,98	1,04	0,91
1QRV	L	0,02	3,71	3,59	3,85	3,86	3,88	3,89	3,89	3,90	3,91	3,91	1,96	3,07	3,22	3,38	3,47	3,54	3,59	3,62	3,73	3,67
1QRV	M	1,20	9,10	8,80	8,78	8,73	8,70	8,68	8,67	8,66	8,66	8,64	9,53	9,29	9,22	9,18	9,15	9,13	9,12	9,11	9,11	9,09
1QRV	V	-0,30	2,43	2,34	3,59	3,73	3,82	3,88	3,92	3,96	3,88	4,00	-2,79	1,70	1,05	1,52	1,80	1,99	2,12	2,22	2,61	2,35
1MNM	K	0,54	347,93	392,94	122,31	94,09	77,15	65,84	57,75	51,68	89,67	43,17	346,39	392,92	119,08	90,70	73,69	62,36	54,28	48,22	87,01	39,75

Capítulo VI: Informação Suplementar

1MNM	E	0,24	-278,00	-322,00	-95,69	-72,90	-59,23	-50,13	-43,63	-38,75	-71,11	-31,91	-282,88	-316,80	-93,16	-69,47	-55,27	-45,82	-39,06	-34,01	-65,48	-26,93
1MNM	R	0,28	313,30	354,47	109,22	83,67	68,34	58,10	50,78	45,28	79,67	37,57	311,54	352,61	106,61	81,00	65,64	55,40	48,10	42,61	77,34	34,94
1MNM	R	0,81	385,14	439,47	143,26	113,02	94,88	82,77	74,12	67,62	108,44	58,50	371,39	428,20	130,74	100,77	82,83	70,90	62,39	56,03	97,68	47,15
1MNM	K	0,64	346,36	400,27	121,95	93,85	76,98	65,71	57,66	51,60	90,65	43,10	349,39	401,77	118,98	90,20	72,95	61,45	53,24	47,09	86,91	38,48
1MNM	I	0,88	-1,12	0,87	0,98	1,24	1,39	1,49	1,55	1,60	1,62	1,67	-6,75	-3,37	-3,70	-3,31	-3,08	-2,92	-2,80	-2,72	-2,49	-2,60
1MNM	E	0,20	-287,02	-330,70	-99,02	-75,53	-61,43	-52,04	-45,32	-40,29	-73,25	-33,23	-287,30	-323,33	-97,34	-73,60	-59,35	-49,85	-43,07	-37,98	-69,70	-30,85
1MNM	I	0,86	4,60	2,40	1,97	1,64	1,43	1,30	1,20	1,12	0,94	1,01	0,78	1,00	1,38	1,46	1,51	1,54	1,57	1,58	1,50	1,61
1MNM	K	-0,13	243,80	282,74	84,02	64,06	52,08	44,10	38,40	34,12	62,71	28,14	239,54	282,49	82,10	62,43	50,63	42,76	37,15	32,93	62,03	27,03
1MNM	F	4,05	1,34	0,18	0,44	0,33	0,27	0,23	0,20	0,17	0,05	0,14	0,17	-0,92	0,23	0,24	0,24	0,25	0,25	0,25	0,00	0,25
1MNM	I	4,05	2,71	2,04	1,21	1,01	0,90	0,82	0,77	0,72	0,74	0,66	0,72	-0,48	-0,23	-0,35	-0,41	-0,46	-0,49	-0,52	-0,66	-0,55
1MNM	E	0,41	-218,57	-256,99	-75,20	-57,29	-46,55	-39,40	-34,29	-30,46	-56,99	-25,10	-225,13	-254,39	-77,68	-59,24	-48,18	-40,79	-35,52	-31,56	-56,75	-26,01
1MNM	N	1,75	-1,32	-3,35	-0,39	-0,27	-0,20	-0,16	-0,13	-0,10	-0,61	-0,07	-1,48	-4,11	0,70	0,96	1,11	1,21	1,29	1,34	0,49	1,41
1MNM	K	-0,13	225,68	258,31	77,57	59,07	47,98	40,58	35,31	31,35	57,37	25,82	217,75	256,83	74,73	56,86	46,14	39,00	33,90	30,07	56,73	24,72
1MNM	R	2,76	246,41	283,71	84,59	64,40	52,30	44,23	38,49	34,18	62,91	28,16	238,33	273,46	80,56	60,86	49,04	41,17	35,55	31,34	59,04	25,45
1MNM	H	-0,81	4,05	1,63	1,76	1,48	1,31	1,19	1,11	1,05	0,85	0,96	0,42	-0,17	-2,02	-2,32	-2,50	-2,61	-2,70	-2,76	-2,57	-2,84
1MNM	V	0,58	-3,27	-0,88	-0,31	0,06	0,29	0,43	0,54	0,62	0,63	0,73	-1,18	0,21	0,73	0,97	1,13	1,23	1,30	1,35	1,32	1,43
1MNM	T	0,64	-3,02	-4,54	-0,66	-0,36	-0,17	-0,05	0,04	0,11	-0,37	0,21	-4,38	-3,93	-0,89	-0,46	-0,20	-0,02	0,10	0,19	-0,09	0,33
1MNM	F	4,05	1,31	0,20	0,47	0,37	0,31	0,27	0,24	0,21	0,09	0,18	1,76	-0,94	0,66	0,52	0,44	0,38	0,34	0,31	-0,16	0,27
1MNM	S	0,43	-1,12	1,02	-0,30	-0,19	-0,13	-0,09	-0,06	-0,04	0,23	-0,01	-0,45	1,69	-0,56	-0,57	-0,58	-0,58	-0,58	-0,58	-0,19	-0,58
1MNM	K	4,05	347,57	411,44	123,35	95,29	78,43	67,17	59,11	53,06	93,45	44,56	344,57	411,09	120,21	92,19	75,39	64,19	56,20	50,20	91,52	41,81
1MNM	R	4,05	389,81	429,99	139,10	107,66	88,75	76,10	67,04	60,22	98,76	50,62	387,78	427,30	135,22	103,62	84,64	71,99	62,94	56,14	95,15	46,62
1MNM	K	2,83	267,91	307,32	92,30	70,36	57,21	48,43	42,17	37,47	68,20	30,91	265,95	309,10	90,77	68,89	55,77	47,02	40,78	36,10	67,58	29,56
1MNM	H	-1,24	5,14	1,62	2,11	1,72	1,48	1,31	1,19	1,10	0,68	0,96	4,49	3,32	1,51	1,14	0,92	0,77	0,67	0,58	0,66	0,47
1MNM	I	3,24	4,17	1,47	1,49	1,16	0,96	0,82	0,73	0,65	0,47	0,55	1,37	-1,38	0,04	-0,13	-0,23	-0,30	-0,34	-0,38	-0,81	-0,43
1MNM	M	4,05	1,05	0,11	0,41	0,34	0,28	0,25	0,23	0,21	0,10	0,18	0,64	-0,52	0,44	0,41	0,39	0,38	0,37	0,36	0,12	0,34
1MNM	K	4,05	359,02	407,64	125,53	96,30	78,73	67,00	58,61	52,30	91,34	43,45	358,04	406,74	122,58	93,16	75,52	63,76	55,37	49,07	88,70	40,26
1MNM	K	4,05	415,10	445,56	145,90	112,26	92,06	78,57	68,94	61,71	101,15	51,56	407,22	444,58	141,57	108,41	88,53	75,29	65,84	58,75	99,41	48,84
1MNM	F	-0,13	0,77	0,56	0,34	0,29	0,25	0,23	0,21	0,20	0,11	0,18	1,45	1,34	0,95	0,88	0,84	0,81	0,80	0,78	0,72	0,76
1MNM	E	2,27	-357,56	-407,99	-123,08	-93,78	-76,20	-64,48	-56,10	-49,81	-89,04	-40,99	-355,85	-404,12	-120,40	-91,02	-73,42	-61,69	-53,33	-47,06	-86,48	-38,29
1MNM	S	0	2,64	2,93	0,98	0,77	0,65	0,56	0,50	0,45	0,66	0,38	2,88	2,91	0,37	0,06	-0,13	-0,25	-0,34	-0,41	-0,16	-0,50
1MNM	V	0,11	6,38	3,21	2,10	1,57	1,26	1,04	0,89	0,78	0,71	0,62	0,55	-0,30	-0,27	-0,37	-0,42	-0,47	-0,49	-0,51	-0,57	-0,54
1MNM	L	4,05	-0,69	-1,58	0,28	0,40	0,48	0,52	0,55	0,57	0,19	0,60	-2,00	-4,29	-0,88	-0,73	-0,64	-0,59	-0,54	-0,52	-1,25	-0,47
1MNM	T	4,05	4,41	3,19	1,54	1,19	0,97	0,83	0,73	0,65	0,75	0,54	3,55	1,37	0,47	0,09	-0,14	-0,29	-0,40	-0,49	-0,58	-0,60

Capítulo VI: Informação Suplementar																						
1MNM	T	0,64	2,24	0,23	0,71	0,52	0,41	0,33	0,28	0,24	0,05	0,18	3,40	0,23	0,51	0,16	-0,06	-0,20	-0,31	-0,38	-0,62	-0,49

Tabela VI.4 – Resultados da energia livre de ligação (kcal mol^{-1}) calculada pelo método ASM, usando estruturas geradas de simulações de DM de 2 ns em solvente explícito, através da resolução da equação linear e não-linear de Poisson-Boltzmann. Constantes dielétricas de 1 a 10.

Proteína	AA Mutado	$\Delta\Delta G_{\text{exp}} / \text{kcal mol}^{-1}$	$\Delta\Delta G_{\text{MM-PBSA}} / \text{kcal mol}^{-1}$																			
			PB linear										PB não-linear									
			$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$	$\epsilon 5$	$\epsilon 6$	$\epsilon 7$	$\epsilon 8$	$\epsilon 9$	$\epsilon 10$
1BDT	M	1,10	3,72	4,09	4,68	4,79	4,85	4,89	4,92	4,93	4,87	4,96	3,55	4,35	4,60	4,73	4,81	4,85	4,89	4,91	4,93	4,94
1BDT	S	1,30	7,41	3,61	2,18	1,52	1,12	0,86	0,67	0,52	0,36	0,32	8,45	4,00	2,52	1,77	1,32	1,02	0,80	0,64	0,51	0,41
1BDT	K	1,30	221,81	251,29	76,46	58,29	47,40	40,13	34,94	31,05	56,03	25,61	-5,57	0,42	2,42	3,43	4,04	4,44	4,74	4,96	5,13	5,27
1BDT	M	1,90	-1,40	-1,74	-0,36	-0,23	-0,15	-0,10	-0,06	-0,03	-0,21	0,01	0,13	0,13	0,13	0,13	0,13	0,13	0,13	0,13	0,13	0,13
1BDT	Q	1,80	5,07	1,75	4,42	4,33	4,28	4,25	4,22	4,20	3,69	4,17	5,84	4,95	4,65	4,50	4,40	4,34	4,30	4,26	4,23	4,21
1BDT	N	2,00	3,85	2,59	3,40	3,35	3,32	3,29	3,28	3,27	3,08	3,26	3,44	3,33	3,28	3,26	3,25	3,24	3,24	3,23	3,23	3,23
1BDT	R	6,00	319,29	365,77	112,64	86,73	71,16	60,76	53,32	47,73	83,29	39,89	-8,19	2,48	5,95	7,71	8,76	9,44	9,92	10,28	10,55	10,76
1BDT	R	0,20	276,42	317,31	95,61	72,98	59,39	50,32	43,84	38,97	70,19	32,14	-10,84	-0,87	2,41	4,03	4,99	5,62	6,06	6,38	6,62	6,81
1BDT	N	-1,00	4,14	4,73	1,42	1,08	0,88	0,75	0,66	0,58	1,04	0,49	-0,32	-0,10	-0,02	0,02	0,04	0,06	0,07	0,08	0,09	0,09
1BDT	R	-0,30	207,84	232,37	71,15	54,08	43,85	37,03	32,17	28,52	51,56	23,43	-1,22	2,00	3,09	3,64	3,99	4,22	4,40	4,53	4,64	4,73
1BDT	S	-2,30	12,41	7,57	4,14	3,10	2,48	2,06	1,77	1,55	1,63	1,24	9,45	4,76	3,20	2,42	1,95	1,63	1,41	1,24	1,10	1,00
1BDT	V	-2,30	9,92	4,57	3,71	2,93	2,47	2,16	1,93	1,77	1,48	1,53	9,15	4,89	3,47	2,76	2,33	2,05	1,84	1,69	1,57	1,47
1BDT	N	3,20	-4,31	-4,03	-1,08	-0,68	-0,44	-0,27	-0,16	-0,07	-0,37	0,05	-3,48	-1,47	-0,81	-0,47	-0,27	-0,14	-0,04	0,03	0,08	0,12
1BDT	S	-0,20	2,08	-0,29	0,90	0,75	0,67	0,62	0,58	0,54	0,29	0,51	2,89	1,59	1,16	0,94	0,82	0,73	0,68	0,63	0,60	0,57
1BDT	Q	-0,40	0,17	-0,36	0,08	0,06	0,06	0,05	0,05	0,04	-0,08	0,04	1,39	0,71	0,48	0,37	0,30	0,25	0,22	0,19	0,17	0,16
1B3T	R	3,41	299,89	342,21	110,52	87,17	73,19	63,87	57,20	52,18	84,56	45,12	-34,05	-3,05	6,65	11,15	13,65	15,20	16,25	16,98	17,51	17,91
1B3T	Y	2,62	11,22	-1,84	8,08	7,70	7,46	7,29	7,15	7,04	5,82	6,88	11,89	9,41	8,51	8,03	7,73	7,51	7,34	7,21	7,11	7,01
1B3T	R	4,40	298,46	315,73	105,60	81,46	66,95	57,24	50,29	45,05	72,46	37,68	14,81	13,90	13,86	13,72	13,55	13,40	13,26	13,13	13,01	12,89
1MSE	S	0,06	3,37	2,55	1,23	0,96	0,80	0,69	0,61	0,55	0,60	0,47	3,71	1,95	1,36	1,07	0,89	0,77	0,68	0,61	0,56	0,52
1MSE	N	0,60	1,03	2,09	0,42	0,34	0,29	0,26	0,23	0,22	0,47	0,19	-1,72	-0,76	-0,45	-0,29	-0,20	-0,14	-0,09	-0,06	-0,04	-0,02
1MSE	E	-0,10	-121,75	-139,63	-41,68	-31,68	-25,69	-21,69	-18,84	-16,70	-30,90	-13,72	5,69	1,28	-0,20	-0,95	-1,40	-1,70	-1,92	-2,09	-2,22	-2,33
1MSE	S	0,10	5,52	3,27	2,42	2,02	1,79	1,63	1,52	1,43	1,35	1,31	6,11	3,52	2,65	2,21	1,94	1,76	1,63	1,53	1,46	1,39
1QRV	L	0,02	3,71	3,59	3,85	3,86	3,88	3,89	3,89	3,90	3,91	3,91	3,90	3,89	3,89	3,90	3,90	3,90	3,91	3,91	3,91	3,91
1QRV	M	1,20	9,10	8,80	8,78	8,73	8,70	8,68	8,67	8,66	8,66	8,64	10,00	9,30	9,02	8,94	8,87	8,82	8,79	8,76	8,74	8,72
1QRV	V	-0,30	2,43	2,34	3,59	3,73	3,82	3,88	3,92	3,96	3,88	4,00	2,21	3,18	3,51	3,67	3,77	3,84	3,89	3,92	3,95	3,98
1MNM	K	0,54	347,93	392,94	122,31	94,09	77,15	65,84	57,75	51,68	89,67	43,17	-12,74	0,51	4,92	7,11	8,41	9,27	9,89	10,34	10,69	10,97
1MNM	E	0,24	-278,00	-322,00	-95,69	-72,90	-59,23	-50,13	-43,63	-38,75	-71,11	-31,91	22,46	7,17	2,05	-0,51	-2,04	-3,07	-3,80	-4,35	-4,78	-5,12

Capítulo VI: Informação Suplementar

1MNM	R	0,28	313,30	354,47	109,22	83,67	68,34	58,10	50,78	45,28	79,67	37,57	-9,84	0,40	3,81	5,51	6,53	7,20	7,68	8,03	8,31	8,52
1MNM	R	0,81	385,14	439,47	143,26	113,02	94,88	82,77	74,12	67,62	108,44	58,50	-20,93	5,12	13,45	17,44	19,73	21,19	22,19	22,92	23,46	23,87
1MNM	K	0,64	346,36	400,27	121,95	93,85	76,98	65,71	57,66	51,60	90,65	43,10	-19,66	-2,83	2,77	5,55	7,20	8,29	9,06	9,63	10,06	10,40
1MNM	I	0,88	-1,12	0,87	0,98	1,24	1,39	1,49	1,55	1,60	1,62	1,67	-2,41	-0,25	0,48	0,85	1,07	1,22	1,32	1,40	1,46	1,51
1MNM	E	0,20	-287,02	-330,70	-99,02	-75,53	-61,43	-52,04	-45,32	-40,29	-73,25	-33,23	15,03	3,16	-0,78	-2,74	-3,93	-4,72	-5,28	-5,70	-6,03	-6,30
1MNM	I	0,86	4,60	2,40	1,97	1,64	1,43	1,30	1,20	1,12	0,94	1,01	3,20	1,83	1,39	1,18	1,05	0,97	0,91	0,86	0,83	0,80
1MNM	K	-0,13	243,80	282,74	84,02	64,06	52,08	44,10	38,40	34,12	62,71	28,14	-15,51	-4,26	-0,50	1,39	2,52	3,28	3,83	4,23	4,56	4,82
1MNM	F	4,05	1,34	0,18	0,44	0,33	0,27	0,23	0,20	0,17	0,05	0,14	0,32	0,16	0,11	0,09	0,07	0,06	0,06	0,05	0,05	0,05
1MNM	I	4,05	2,71	2,04	1,21	1,01	0,90	0,82	0,77	0,72	0,74	0,66	1,69	1,01	0,79	0,69	0,62	0,58	0,56	0,54	0,52	0,51
1MNM	E	0,41	-218,57	-256,99	-75,20	-57,29	-46,55	-39,40	-34,29	-30,46	-56,99	-25,10	16,48	5,17	1,38	-0,52	-1,67	-2,44	-5,04	-3,41	-3,74	-4,01
1MNM	N	1,75	-1,32	-3,35	-0,39	-0,27	-0,20	-0,16	-0,13	-0,10	-0,61	-0,07	1,75	0,92	0,63	0,49	0,40	0,34	1,12	0,26	0,24	0,22
1MNM	K	-0,13	225,68	258,31	77,57	59,07	47,98	40,58	35,31	31,35	57,37	25,82	-7,25	-0,59	1,65	2,78	3,47	3,93	4,26	4,52	4,72	4,88
1MNM	R	2,76	246,41	283,71	84,59	64,40	52,30	44,23	38,49	34,18	62,91	28,16	-10,90	-2,24	0,68	2,17	3,08	3,70	4,15	4,50	4,77	4,99
1MNM	H	-0,81	4,05	1,63	1,76	1,48	1,31	1,19	1,11	1,05	0,85	0,96	3,96	2,28	1,71	0,42	1,26	1,15	1,08	1,02	0,97	0,93
1MNM	V	0,58	-3,27	-0,88	-0,31	0,06	0,29	0,43	0,54	0,62	0,63	0,73	-3,18	0,64	-0,30	0,06	0,28	0,43	0,53	0,61	0,67	0,72
1MNM	T	0,64	-3,02	-4,54	-0,66	-0,36	-0,17	-0,05	0,04	0,11	-0,37	0,21	-0,26	0,06	0,18	0,25	0,30	0,33	0,36	0,38	0,40	0,42
1MNM	F	4,05	1,31	0,20	0,47	0,37	0,31	0,27	0,24	0,21	0,09	0,18	1,16	0,61	0,42	0,33	0,27	0,23	0,21	0,19	0,17	0,16
1MNM	S	0,43	-1,12	1,02	-0,30	-0,19	-0,13	-0,09	-0,06	-0,04	0,23	-0,01	-3,25	-1,55	-0,98	-0,70	-0,53	-0,42	-0,34	-0,28	-0,23	-0,19
1MNM	K	4,05	347,57	411,44	123,35	95,29	78,43	67,17	59,11	53,06	93,45	44,56	-12,83	1,36	6,03	8,36	9,75	10,65	11,29	11,75	12,11	12,39
1MNM	R	4,05	389,81	429,99	139,10	107,66	88,75	76,10	67,04	60,22	98,76	50,62	2,87	14,82	16,69	17,39	17,69	17,80	17,83	17,82	17,78	16,43
1MNM	K	2,83	267,91	307,32	92,30	70,36	57,21	48,43	42,17	37,47	68,20	30,91	-13,75	-3,11	0,47	2,27	3,35	4,08	5,42	5,01	5,33	5,58
1MNM	H	-1,24	5,14	1,62	2,11	1,72	1,48	1,31	1,19	1,10	0,68	0,96	6,05	3,21	-1,53	1,79	1,50	1,31	1,18	1,07	0,99	0,93
1MNM	I	3,24	4,17	1,47	1,49	1,16	0,96	0,82	0,73	0,65	0,47	0,55	4,33	2,24	1,55	1,21	1,00	0,86	0,76	0,68	0,62	0,57
1MNM	M	4,05	1,05	0,11	0,41	0,34	0,28	0,25	0,23	0,21	0,10	0,18	1,53	0,81	0,56	0,44	0,36	0,31	0,27	0,24	0,22	0,20
1MNM	K	4,05	359,02	407,64	125,53	96,30	78,73	67,00	58,61	52,30	91,34	43,45	-16,11	-1,51	3,31	5,72	7,14	8,07	8,73	9,21	9,58	9,87
1MNM	K	4,05	415,10	445,56	145,90	112,26	92,06	78,57	68,94	61,71	101,15	51,56	-20,30	0,71	7,21	10,23	11,91	12,94	13,62	14,10	14,44	14,70
1MNM	F	-0,13	0,77	0,56	0,34	0,29	0,25	0,23	0,21	0,20	0,11	0,18	0,37	0,22	0,17	0,15	0,13	0,12	0,11	0,11	0,10	0,10
1MNM	E	2,27	-357,56	-407,99	-123,08	-93,78	-76,20	-64,48	-56,10	-49,81	-89,04	-40,99	12,96	0,92	-3,02	-4,95	-6,09	-6,84	-7,36	-7,75	-8,04	-8,27
1MNM	S	0	2,64	2,93	0,98	0,77	0,65	0,56	0,50	0,45	0,66	0,38	0,30	0,27	0,26	0,24	0,24	0,23	0,22	0,21	0,20	0,20
1MNM	V	0,11	6,38	3,21	2,10	1,57	1,26	1,04	0,89	0,78	0,71	0,62	5,82	2,90	1,93	1,44	1,15	0,96	0,82	0,72	0,64	0,58
1MNM	L	4,05	-0,69	-1,58	0,28	0,40	0,48	0,52	0,55	0,57	0,19	0,60	-2,18	-0,70	-0,21	0,03	0,17	0,27	0,33	0,38	0,42	0,45
1MNM	T	4,05	4,41	3,19	1,54	1,19	0,97	0,83	0,73	0,65	0,75	0,54	1,66	0,92	0,67	0,54	0,47	0,42	0,38	0,35	0,33	0,31
1MNM	T	0,64	2,24	0,23	0,71	0,52	0,41	0,33	0,28	0,24	0,05	0,18	1,84	0,91	0,60	0,45	0,36	0,29	0,25	0,22	0,19	0,17

Tabela VI.5 - Resultados dos testes estatísticos para as 60 mutações.

Global										
		VP	FP	FN	VN	P	R	F1	A	E
$\Delta\Delta G_{MM-PBSA}$	$\epsilon 1$	6	19	14	21	24	30	27	45	53
	$\epsilon 2$	6	17	14	23	26	30	28	48	58
	$\epsilon 3$	9	17	11	23	35	45	39	53	58
	$\epsilon 4$	11	16	9	24	41	55	47	58	60
	$\epsilon 5$	11	15	9	25	42	55	48	60	63
	$\epsilon 6$	11	15	9	25	42	55	48	60	63
	$\epsilon 7$	11	14	9	26	44	55	49	62	65
	$\epsilon 8$	11	14	9	26	44	55	49	62	65
	$\epsilon 9$	11	14	9	26	44	55	49	62	65
	$\epsilon 10$	11	14	9	26	44	55	49	62	65
$\Delta\Delta G_{ele} + \Delta\Delta G_{solv\ polar}$	$\epsilon 1$	4	13	16	27	24	20	22	52	68
	$\epsilon 2$	4	9	16	31	31	20	24	58	78
	$\epsilon 3$	7	8	13	32	47	35	40	65	80
	$\epsilon 4$	9	11	11	29	45	45	45	63	73
	$\epsilon 5$	9	11	11	29	45	45	45	63	73
	$\epsilon 6$	9	9	11	31	50	45	47	67	78
	$\epsilon 7$	9	9	11	31	50	45	47	67	78
	$\epsilon 8$	9	9	11	31	50	45	47	67	78
	$\epsilon 9$	9	9	11	31	50	45	47	67	78
	$\epsilon 10$	9	9	11	31	50	45	47	67	78

Tabela VI.6 - Resultados dos testes estatísticos para os resíduos carregados.

Carregados										
		VP	FP	FN	VN	P	R	F1	A	E
$\Delta\Delta G_{MM-PBSA}$	$\epsilon 1$	3	6	7	9	33	30	32	48	60
	$\epsilon 2$	3	7	7	8	30	30	30	44	53
	$\epsilon 3$	7	8	3	7	47	70	56	56	47
	$\epsilon 4$	9	8	1	7	53	90	67	64	47
	$\epsilon 5$	9	9	1	6	50	90	64	60	40
	$\epsilon 6$	9	9	1	6	50	90	64	60	40
	$\epsilon 7$	9	9	1	6	50	90	64	60	40
	$\epsilon 8$	9	9	1	6	50	90	64	60	40
	$\epsilon 9$	9	9	1	6	50	90	64	60	40
	$\epsilon 10$	9	9	1	6	50	90	64	60	40
$\Delta\Delta G_{ele} + \Delta\Delta G_{solv\ polar}$	$\epsilon 1$	2	5	8	10	29	20	24	48	67
	$\epsilon 2$	2	4	8	11	33	20	25	52	73
	$\epsilon 3$	6	5	4	10	55	60	57	64	67
	$\epsilon 4$	8	8	2	7	50	80	62	60	47
	$\epsilon 5$	9	9	1	6	50	90	64	60	40
	$\epsilon 6$	9	9	1	6	50	90	64	60	40

Capítulo VI: Informação Suplementar

$\epsilon 7$	9	9	1	6	50	90	64	60	40
$\epsilon 8$	9	9	1	6	50	90	64	60	40
$\epsilon 9$	9	9	1	6	50	90	64	60	40
$\epsilon 10$	9	9	1	6	50	90	64	60	40

Tabela VI.7 - Resultados dos testes estatísticos para os resíduos polares.

Polar										
	VP	FP	FN	VN	P	R	F1	A	E	
$\Delta\Delta G_{MM-PBSA}$	$\epsilon 1$	2	6	2	8	25	50	33	56	57
	$\epsilon 2$	2	4	2	10	33	50	40	67	71
	$\epsilon 3$	2	4	2	10	33	50	40	67	71
	$\epsilon 4$	2	3	2	11	40	50	44	72	79
	$\epsilon 5$	2	1	2	13	67	50	57	83	93
	$\epsilon 6$	2	1	2	13	67	50	57	83	93
	$\epsilon 7$	2	1	2	13	67	50	57	83	93
	$\epsilon 8$	2	1	2	13	67	50	57	83	93
	$\epsilon 9$	2	1	2	13	67	50	57	83	93
	$\epsilon 10$	2	1	2	13	67	50	57	83	93
$\Delta\Delta G_{ele} + \Delta\Delta G_{solv\ apolar}$	$\epsilon 1$	1	5	3	9	17	25	20	56	64
	$\epsilon 2$	1	3	3	11	25	25	25	67	79
	$\epsilon 3$	1	2	3	12	33	25	29	72	86
	$\epsilon 4$	1	2	3	12	33	25	29	72	86
	$\epsilon 5$	0	2	4	12	0	0	0	67	86
	$\epsilon 6$	0	0	4	14	0	0	0	78	100
	$\epsilon 7$	0	0	4	14	0	0	0	78	100
	$\epsilon 8$	0	0	4	14	0	0	0	78	100
	$\epsilon 9$	0	0	4	14	0	0	0	78	100
	$\epsilon 10$	0	0	4	14	0	0	0	78	100

Tabela VI.8 - Resultados dos testes estatísticos para os resíduos apolares.

Apolar										
		VP	FP	FN	VN	P	R	F1	A	E
$\Delta\Delta G_{MM-PBSA}$	$\epsilon 1$	1	7	5	4	13	17	14	29	36
	$\epsilon 2$	1	6	5	5	14	17	15	35	45
	$\epsilon 3$	0	5	6	6	0	0	0	35	55
	$\epsilon 4$	0	5	6	6	0	0	0	35	55
	$\epsilon 5$	0	5	6	6	0	0	0	35	55
	$\epsilon 6$	0	5	6	6	0	0	0	35	55
	$\epsilon 7$	0	4	6	7	0	0	0	41	64
	$\epsilon 8$	0	4	6	7	0	0	0	41	64
	$\epsilon 9$	0	4	6	7	0	0	0	41	64
	$\epsilon 10$	0	4	6	7	0	0	0	41	64
$\Delta\Delta G_{ele} + \Delta\Delta G_{solv\ apolar}$	$\epsilon 1$	1	3	5	8	25	17	20	53	73
	$\epsilon 2$	1	2	5	9	33	17	22	59	82
	$\epsilon 3$	0	1	6	10	0	0	0	59	91
	$\epsilon 4$	0	1	6	10	0	0	0	59	91
	$\epsilon 5$	0	0	6	11	0	0	0	65	100
	$\epsilon 6$	0	0	6	11	0	0	0	65	100
	$\epsilon 7$	0	0	6	11	0	0	0	65	100
	$\epsilon 8$	0	0	6	11	0	0	0	65	100
	$\epsilon 9$	0	0	6	11	0	0	0	65	100
	$\epsilon 10$	0	0	6	11	0	0	0	65	100